Investigating Components of Prospective Cognition in Gibbons (*Hylobatidae*)

A thesis submitted for the degree of Doctor of Philosophy (Ph.D.)

by

Heather K Taylor

Social and Health Sciences
Abertay University.

February, 2019

Declaration

Candidate's declarations:

I, Heather Taylor, hereby certify that this thesis submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy (PhD), Abertay University, is wholly my own work unless otherwise referenced or acknowledged. This work has not been submitted for any other qualification at any other academic institution.

Signed ……[redacted] ……………………………………………………

Date…26.2.19..…………………………………………………………………..

Supervisor's declaration:

I, Dr. Clare Cunningham, hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy (PhD) in Abertay University and that the candidate is qualified to submit this thesis in application for that degree.

Signed…… [redacted] ……..…….…………...

Date…26.02.2019………………………………………………….……….……..

Certificate of Approval

I certify that this is a true and accurate version of the thesis approved by the examiners, and that all relevant ordinance regulations have been fulfilled.

Supervisor… [redacted] ………………………………..…….

Date……19/09/2019………………………………………………………………..

Acknowledgements

First, I would like to express my sincere gratitude to my primary supervisor, Dr. Clare Cunningham. I don't think I have ever told her, but a short 10 minute presentation she gave in 2011 made me finally decide I definitely wanted to do 'something with animals and psychology'. She has been a consistent source of guidance, friendship and support since, and I would not be where I am today without her. I also thank my second supervisor, Dr. Scott Hardie, for his helpful comments, support, and allowing me the opportunity to explore my interest in gibbon handedness. Thanks also go to my examiners, Dr. Lara Wood and Dr. Susan Cheyne, for reviewing and providing helpful comments on my thesis, and (perhaps more so!), for making my viva an enjoyable experience.

Thank you to the staff at the Gibbon Conservation Center, Gabi, Melissa, Tiffany, Alma and Jesse. I am very grateful for the opportunity to conduct research at the centre, for your help with all of my requests, the many (many) trips to buy blueberries, and for sharing your American-isms, home and work place. You all inspire me in different ways. Thanks also go to Chris, Carmen, Valerie, Brittany, Neta, and all the other volunteers and individuals associated with the gibbon center for always being welcoming and kind, it really was appreciated when I was so far from home.

I thank my parents for their continued support and encouragement over the past however many years I have been studying. In particular, thanks go to my mum for proof-reading and for always being available to listen to me ramble about gibbons. To my dad, I sincerely appreciate your continued interest in what I do, and for helping me to build the apparatus needed to do it. More importantly, I thank you both for your unwavering faith in my abilities.

To all the friends who encouraged me when I doubted myself, made me laugh when I needed it the most, and reminded me there is more to life than a Ph.D., thank you.

Finally, I would like to acknowledge the most important contributors to this thesis. For their hard-work, motivation, and ability to make even the arduous task of video coding (mostly!) enjoyable, I thank the gibbons.

Abstract

Complex prospection, requiring projection of one's mental self into a subjective future time, has been proposed to be uniquely human, with ongoing debate as to the criteria necessary for assessment of this ability. Focus on this dichotomous framework of future directed thinking, assessing whether non-human primates possess levels of prospection comparable to that displayed by humans, has detracted from assessment of the range of abilities and species specific adaptations found across primates. Current evidence for advanced prospection within non-human primates remains contested; however, the large apes have displayed greater proficiency at tasks requiring prospective cognition than monkeys. This thesis investigates components of prospective cognition within the small apes (*Hylobatidae*). Gibbons represent the only surviving divergence between large apes and monkeys. As such, assessment of basic prospection and component processes within gibbons allows for a clearer overview of the emergence of prospective abilities across the primate order.

Here, 31 gibbons (*Hoolock leuconedys*, n = 9, *Hylobates moloch*, n = 9, *Hylobates pileatus*, n = 5, *Nomascus leucogenys*, n = 6, *Symphalangus syndactylus*, n = 2) were first assessed on their ability to attend to the functionally relevant features of two rakes. One functional and one non-functional rake were presented during a raking in task, requiring selection of a functional rake in order to draw in an out of reach food reward. Pilot testing provided little support for this ability; however, given further testing with rake sets presenting more distinctive perceptual differences, gibbons were found to reliably distinguish between functional and non-functional rakes. Some evidence was found for subjects transferring knowledge across different rake sets, with subjects reaching criterion level performance faster during later experiments.

Once gibbons had learned the necessary skills to select a functional rake for reward retrieval, a series of experiments assessed their capacity for basic prospection. Subjects (*H. leuconedys*, n = 5, *N.*

leucogenys, n = 3, *H. pileatus*, n = 3, *S. syndactylus*, n = 1) were again required to select between one functional and one non-functional rake; however, these rakes were now un-baited. This allowed for assessment of whether gibbons would select functional rakes for future use, with the selected rake being baited at a separate location following a time delay. Subjects reliably selected the functional rake when delays of up to five minutes were imposed between rake selection and reward retrieval. The increasing time delay did not greatly affect subjects' performance, suggesting gibbons can relate temporally and spatially distinct events, displaying basic prospection.

Finally, a preliminary investigation of self-control capacities within gibbons (*H. leuconedys*, n = 3, *H. pileatus*, n = 3, *N. leucogenys*, n = 2) was conducted. Self-control is an integral feature of much prospective behaviour without which an individual cannot inhibit current desires in favour of future ones. Individual differences were found; however, three gibbons refrained from selecting an immediately attainable small reward, instead selecting a rake functional for retrieval of a larger reward at a second location. Taken together, the current findings provide initial evidence of both basic prospection and self-control within the small apes.

Table of Contents

List of Figures

List of Tables

Chapter 1: Taxonomy and physical cognition within the small apes (*Hylobatidae*)

1.1 Introduction

Gibbons, the small apes of the family *Hylobatidae*, display unique morphological, behavioural and genetic characteristics, representing alternative solutions to lineage survival compared to those present in large apes (Reichard, Barelli, Hirai & Nowak, 2016; Thinh et al., 2010a). Found across South, East, and Southeast Asia, *Hylobatidae* are the most specious ape taxa, offering an opportunity for cross-species comparisons of cognition and behaviour, and assessment of species specific adaptations to ecological demands. Despite this, gibbons have been largely neglected within scientific research, with early assessments reporting low motivation and disinterest in experimental testing within this group (see section 1.4.1). However, more recent investigations incorporating tasks better suited to gibbons' morphology have found performance levels more representative of gibbons' phylogenetic position, located between the large apes and monkeys (see sections 1.4.2 and 1.4.3). The main aim of this thesis is to first assess whether gibbons can attend to the functionally relevant features of two rakes during a raking in task, before assessing components of prospective cognition. An assessment of basic prospection, selecting functional rakes for future use, is given, followed by a preliminary investigation into self-control capacities within the small apes. To set the overall context, this chapter first presents an overview of gibbon taxonomy, before a brief discussion of ecology and social systems within the small apes. A review of relevant cognitive research is then presented, with a focus on assessments of object manipulation and tool use within the small apes.

1.2 Taxonomy

1.2.1 Genera

Gibbons began their independent evolution from the main hominoid line ~19 million years ago (mya) with a subsequent radiation resulting in the emergence of the four main genera; *Nomascus*, *Symphalangus*, *Hoolock* and *Hylobates* (Chatterjee, 2006; Geissmann, 2002a; Harrison, 2016; Mootnick & Groves, 2005; Reichard et al., 2016; Roos, 2016; Roos & Geissmann, 2001; Takacs, Morales, Geissmann & Melnick, 2005; Thinh et al., 2010a). These genera diversified rapidly, with Reichard et al. (2016) noting the genetic differences among hylobatids exceed the range of genetic variation found between other hominoids. The molecular distances between the four gibbon genera are comparable with, or greater than, those between *Homo* and *Pan*, facilitating cross genera comparisons (Roos & Geissmann, 2001). Potentially due in part to this rapid radiation, the estimated divergence dates of the extant gibbon genera remain debated, with the following dates given from Harrison (2016), who presents averages from previous literature. The first extant genus to diverge is proposed to be *Nomascus*, at ~8 mya, followed by *Symphalangus* at ~6.7 mya, with *Hoolock* and *Hylobates* diverging ~6.2mya (Harrison, 2016). Following divergence from the main hylobatid line, *Nomascus* species began to diversify at ~4.24 mya, with *Hylobates* species diversifying between ~3 and ~3.9 mya, and *Hoolock* species between ~1.3 and ~1.8 mya (Reichard et al., 2016; Thinh et al., 2010a).

The four genera differ significantly from each other, most notably in terms of chromosome numbers (diploid chromosomes: *Hoolock* = 38, *Hylobates* = 44, *Nomascus* = 52, *Symphalangus* = 50), but also in molecular, acoustic and morphological features that may have evolved due to ecological niche separation of the genera (Reichard et al., 2016; Roos, 2016). Twenty gibbon species are currently recognised (see Table 1.1), however, there is ongoing debate as to the classification of, and phylogenetic relationships between, these species and subspecies (Brandon-Jones et al., 2004; Chatterjee, 2006, 2009;

Geissmann 1995, 2002a; Harrison, 2016; Reichard et al., 2016; Roos, 2016; Roos & Geissmann, 2001; Takacs et al., 2005; Thinh et al., 2010a).

Table 1.1: Classification of gibbons (*Hylobatidae*) currently recognised. Adapted from Roos (2016).

Genus	Species	Subspecies	Common Name
Nomascus	*N. annamensis*		Northern yellow cheeked gibbon
	N. concolor		Western black crested gibbon
		N. c. concolor	Tonkin black crested gibbon
		N. c. lu	Laotian black crested gibbon
	N. gabriellae		Southern yellow cheeked gibbon
	N. hainanus		Hainan crested gibbon
	N. leucogenys		Northern white cheeked gibbon
	N. nasutus		Eastern black crested gibbon
	N. siki		Southern white cheeked gibbon
Symphalangus	*S. syndactylus*		Siamang
Hoolock	*H. hoolock*		Western hoolock gibbon
	H. leuconedys		Eastern hoolock gibbon
	H. tianxing		Skywalker hoolock gibbon
Hylobates	*H. abbotti*		Abott’s grey gibbon
	H. agilis		Agile gibbon

	H. albibarbis		Bornean white-bearded gibbon
	H. funereus		East Bornean grey gibbon
	H. klossii		Kloss's gibbon
	H. lar		Lar gibbon
		H. l. carpenteri	Carpenter's lar gibbon
		H. l. entelloides	Central lar gibbon
		H. l. lar	Malayan lar gibbon
		H. l. vestitus	Sumatran lar gibbon
		H. l. yunnanensis	Yunnan lar gibbon
	H. moloch		Javan gibbon
	H. muelleri		Müller's gibbon
	H. pileatus		Pileated gibbon

1.2.1.1 *Nomascus*

There is continuing discussion regarding the species and subspecies of the *Nomascus* genus; seven species, with two subspecies of *N. concolor*, are currently recognised (Reichard et al., 2016; Roos, 2016; Thinh et al., 2010a, 2010c; Thinh, Mootnick, Thanh, Nadler & Roos, 2010b). Having the second largest average body weight of the four genera, *Nomascus* species weigh between 7-8 kg, with little body mass dimorphism (males average 1.7% larger than females) (Reichard et al., 2016). All *Nomascus* species show sexual dichromatism, with males having black pelage and females showing various shades of beige, yellow or orange between species (Mootnick & Fan, 2011; Reichard et al., 2016; Thinh et al., 2010c; Tuttle, 2014).

The northern (*N. leucogenys*), and southern (*N. siki*) white cheeked gibbons, from Laos, Vietnam and China, are currently considered to be two separate species (Geissmann, 2007; Harding, 2012; Mootnick & Fan, 2011; Reichard et al., 2016; Roos, 2016; Thinh et al., 2010a, b, c). Likewise, the recently classified northern (*N. annamensis*) and the southern (*N. gabriellae*) yellow-cheeked gibbons found in Laos, Vietnam and Cambodia are now considered distinct species (Brandon-Jones et al., 2004; Reichard et al., 2016; Roos, 2016; Thinh et al., 2010a, b, c). The black crested gibbons, found in China, Vietnam and Laos are categorised into eastern (*N. nasutus*), and western (*N. concolor*) species (Brandon-Jones et al., 2004; Geissmann, 2007; Reichard et al., 2016; Roos, 2016, Thinh et al., 2010a, b, c). The number of *N. concolor* subspecies has been disputed. Four subspecies have previously been suggested based on differences in fur length (see Brandon-Jones et al., 2004; Chatterjee, 2009; Geissmann, 2007); however, as morphological and acoustic data do not support this division, two subspecies are now recognised; *N. c. concolor* (Tonkin black crested gibbon) and *N. c. lu* (Laotian black crested gibbon) (Mootnick & Fan, 2011; Reichard et al., 2016; Roos, 2016; Thinh et al., 2010a, c). The Hainan crested gibbon (*N. hainanus*), endemic to Hainan Island (China), was proposed by Brandon-Jones et al. (2004) to be a sub-species of *N. nasutus*, however various sources consider this to be an independent species based on acoustic and genetic data (Geissmann, 2007; Groves, 2001; Mootnick & Fan, 2011; Reichard et al., 2016; Roos, 2016; Thinh et al., 2010a, b, c).

1.2.1.2 *Symphalangus*

Siamang (*S. syndactylus*) is the only species recognised in the genus *Symphalangus*, being found in Malaysia, Thailand and Indonesia (Brandon-Jones et al., 2004; Roos, 2016). The siamang is the largest of the gibbons, weighing between 10-11kg; this genus also shows the most pronounced body mass dimorphism, with males 10.1% heavier than females (Reichard et al., 2016). Siamangs are monochromatic, with both males and females having black pelage (Tuttle, 2014). Both sexes commonly have webbing between the second and third toes, and a large inflatable throat sack, allowing them to amplify their calls (Mootnick, 2006; Reichard et al., 2016). Some authors propose two sub-species of siamang, the Sumatran (*S. s. syndactylus*), and Malayan siamang (*S. s. continentis*), however further research is needed to substantiate this suggestion (Brandon-Jones et al., 2004; Mootnick, 2006; Roos, 2016).

1.2.1.3 *Hoolock*

Hoolock gibbons were previously considered members of the genus *Bunopithecus*, however, *Bunopithecus* refers to a fossil no longer considered related to extant hoolock gibbons (Mootnick & Groves, 2005). These species were therefore grouped into a new genus, *Hoolock* (Mootnick & Groves, 2005; Roos, 2016). There are currently three recognised species of *Hoolock* gibbon (Chan, Mak, Yang & Huang, 2017; Fan et al., 2016). The western (*H. hoolock*) and eastern (*H. leuconedys*) hoolock gibbons were previously categorised as subspecies, but have been reclassified as species (Geissmann, 2007; Mootnick & Groves, 2005). The western hoolock is found in Bangladesh, Myanmar and India, with eastern hoolock populations in Myanmar, India and China (Brandon-Jones et al., 2004; Chetry, Chetry, Ghosh & Singh, 2010). Choudhury (2013) proposed a subspecies of the western hoolock, the Mishmi Hills hoolock (*H. h. mishmiensis*) located in India; however, further data is needed in order to confirm the status of this subspecies. The recently identified third species, the skywalker hoolock (*H. tianxing*), is found in China and Myanmar (Fan et al., 2016). Hoolock gibbons weigh between 6-7 kg, and show minimal body mass dimorphism but marked sexual dichromatism (Mootnick 2006; Reichard et al., 2016). Adult males in all species are black with white brows and females buff or tan coloured with a white face ring.

1.2.1.4 *Hylobates*

The genus *Hylobates* is the most specious genus, with ongoing debate regarding recognised species and subspecies (Roos, 2016). Gibbons in this genus have a weight range of 5-6 kg with body mass dimorphism evident; males are on average 5.6% heavier than females (Reichard et al., 2016). The following species information is taken from a recent review of gibbon phylogeny presented by Roos (2016), with nine recognised species and five subspecies of *H. lar*.

Kloss's gibbons (*H. klossii*), found in Indonesia, are the only monochromatic species in this genus, with both sexes having black pelage

(Brandon-Jones et al., 2004; Mootnick, 2006; Tuttle, 2014). Pileated gibbons (*H. pileatus*) are found in Cambodia, Laos and Thailand, with this species showing sexual dichromatism; the females have silvery-grey or fawn-buff coats, and the males are black (Brandon-Jones et al., 2004; Mootnick 2006; Reichard et al., 2016; Tuttle, 2014). The Javan gibbons (*H. moloch*), endemic to the island of Java (Indonesia) also display slight sexual dimorphism; both sexes are silver grey however females have darker caps and chests (Tuttle, 2014). Two subspecies of *H. moloch* have been proposed, *H. moloch moloch* and *H moloch pongoalsoni*; however, further genetic research is needed to confirm this separation (Andayani, Morales, Forstner, Supriatna & Melnick, 2001; Brandon-Jones et al., 2004; Dallman & Geissmann 2009; Mootnick, 2006). Müller's gibbon (*H. muelleri*) is found in Indonesia, with this species being polychromatic; the coat colour of both sexes range from shades of brown to grey (Brandon-Jones et al., 2004; Tuttle, 2014). *H. abbotti* (Abott's grey gibbon) and *H. funereus* (East Bornean grey gibbon), both found in Indonesia and Malaysia, have previously been proposed as subspecies of *H. muelleri* (Brandon-Jones et al., 2004; Geissmann, 2002a, 2007; Groves, 2001; Mootnick, 2006). However, distinct genetic and morphological differences suggest these are distinct species (Roos, 2016, see also, Reichard et al., 2016, Thinh et al., 2010a)

H. albibarbis (Bornean white-bearded gibbon), found in Indonesia, has previously been proposed to be a subspecies of *H. agilis* (agile gibbon) found in Indonesia, Malaysia and Thailand (Brandon-Jones et al., 2004; Geissmann, 2002a). However, these groups are recognised as distinct species by various other sources (Chatterjee, 2009; Geissmann, 2007; Groves, 2001; Mootnick, 2006; Reichard et al., 2016; Roos, 2016; Thinh et al., 2010a). Further research is needed to assess the validity of a further proposed subspecies of *H. agilis*, *H. a. unko* (lowland agile gibbon) (Brandon-Jones et al., 2004; Mootnick, 2006; Roos, 2016). The lar, or white-handed gibbon, (*H. lar*) is found across Indonesia, Laos, Malaysia, Myanmar, Thailand, and China. Both lar and agile gibbons are considered asexually dichromatic by Tuttle (2014), with the male and female ranging from cream to brown or black. Five lar sub-species are recognised; *H. l. lar* (Malayan lar gibbon), *H. l. carpenteri* (Carpenter's lar gibbon), *H. l. entelloides* (central lar gibbon), *H. l. vestitus* (Sumatran lar gibbon), and *H. l. yunnanensis* (Yunnan lar gibbon) (Brandon-Jones et al., 2004; Geissmann, 2007; Groves, 2001; Mootnick, 2006; Reichard et al., 2016; Roos, 2016; Thinh et al., 2010a).

1.3 Ecology and social systems

1.3.1 Home range and territoriality

Inhabiting lowland and montane forest across South, East, and Southeast Asia, gibbons are diurnal and arboreal (Harding, 2012; Reichard et al., 2016). Habitat destruction and fragmentation through deforestation and farming pose major threats to the future of gibbon populations, along with continued illegal hunting for the pet trade and medicinal use (Geissmann, 2007). Most gibbon species are considered endangered by the IUCN red list (IUCN, 2018), with four species considered critically endangered (*N. nasutus*, *N. leucogenys*, *N. hainanus*, *N. concolor*) and the eastern hoolock (*H. leuconedys*) considered vulnerable. Malone and Fuentes (2009) estimate the range occupied by gibbon species to be 4000 km (in a north-south direction), with varying habitats both within and between species. Species are separated by rivers in many locations; however, there are several contact areas with groups often living in close proximity, or sympatrically with shared habitats (Chatterjee, 2009; Reichard et al., 2016). The Siamang is the only gibbon known to be sympatric with other species across the whole of its range, sharing territories with *H. agilis* in southern Sumatra, and *H. lar* in northern Sumatra and Malay Peninsula (Chatterjee, 2009). There is evidence of hybrid zones in contact areas between some closely related species; for example, Mootnick (2006) describes *N. gabriellae* x *N. siki* hybrids occurring in Vietnam, Mootnick and Groves (2005) note *Hoolock* species hybrids occur in Myanmar, and Suwanvecho and Brockelman (2012) report *H. lar* x *H. pileatus* hybrids in Thailand.

Gibbons spend the majority of their time in high canopy, with Channa (2007, as cited in Harding, 2012) reporting *N. gabriellae* spend 82.37% of time in high canopy, compared to 17.14% in medium canopy, and just 0.5% in low canopy levels. Gibbons primary mode of locomotion, used during 50-80% of their traveling time, is brachiation, defined by Reichard et al. (2016) as using the forelimbs in an upright suspensory fashion for locomotion (see also, Michilsens, Vereecke, D'Aout & Aerts, 2009). Gibbons are active for around nine hours per day, spending the greatest percentage of their daily time budget resting, followed by feeding, travel, socialising and then other activities (Fan, Ai, Fei, Zhang & Yuan, 2013; Harding, 2012; Phoonjampa, Koenig, Borries & Gale, 2010).

Considered to be frugivorous, gibbons' diets consist primarily of fruit, and in particular figs, with leaves, flowers and insects also consumed (Elder, 2009). Variations in diet occur both at inter-species and intra-species level, with ecological variation in available resources resulting in dietary flexibility (Elder, 2009; Fan et al., 2013; Malone & Fuentes, 2009). Zhang et al. (2014, see also, Bartlett, 2009; Fan et al., 2013) suggest gibbons' ranging behaviour may be determined by food availability and distribution. Gibbons have been found to utilise different areas of their total home range at different times throughout the year dependent on resource availability (Zhang et al., 2014). Furthermore, reports of daily path length have been found to correlate positively with fruit abundance, with average distances travelled decreasing when fruit is scarce (Bartlett, 2009, Zhang et al., 2014). Chivers (1984) reports an average home range across species of approximately 35-50ha. However, more recent reports have suggested smaller ranges, with habitat fragmentation resulting in decreased home range size. For example, Sarma and Kumar (2016) assessed home ranges of four *H. leuconedys* groups, reporting home ranges varying from 1.09-2.49ha within the two groups living in fragmented forest, compared to ranges of 16.28-25.63ha for the two groups living in continuous forest.

Both inter- and intra-specific territorial behaviours are reported in gibbons (Suwanvecho & Brockleman, 2012). For example, Sarma and Kumar (2016) suggest *Hoolock* species defend between 77-95% of their home range as an exclusive territory, with Kim, Lappan and Choe (2011) proposing an area of exclusive use ranging between 67-81% for three *H. moloch* groups. All gibbon species produce a range of structurally complex vocalisations that are species, and in some cases, sex-specific (Geissmann, 2002b). These calls can be solo or in pair duets, with duetting occurring in all but two species (*H. moloch*, *H. klossii*) (Geissmann, 2002b; Koda, 2016). Geissmann suggests these two species may present an adaptation from the common ancestor of all current hylobatid species, which did produce a duet, with the contributions of each sex to the duet becoming increasingly independent over time (see Geissmann, 2002b for further discussion). As well as potentially strengthening pair bonds within groups, calls are proposed to play a role in inter-group communication, mate attraction and territorial defence (Geissmann, 2002b; Koda, 2016).

1.3.2 Social and mating systems

Recent findings suggest much greater flexibility in gibbon's social and mating systems than the previously reported static, pair-bonded family units (Malone & Fuentes, 2009; Morino, 2009; Whittaker & Lappan, 2009). Analysis of the genetic and social ties between groups has developed a better understanding of the complex and dynamic relationships within these communities. Although family groups, consisting of a breeding pair and their offspring are common, individuals may live in groups with multiple adults of the same sex, or even alone in a semi-solitary manner (Malone & Fuentes, 2009; Reichard, Ganpanakngan & Barelli, 2012; Whittaker & Lappan, 2009). In contrast to the monogamous mating system suggested by initial observations, long-term and genetic research has revealed more variable mating patterns in gibbon populations, with multiple reports of polygamy, extra-pair copulations and extra-pair paternity (Barelli et al., 2013; see Morino, 2009 for a review).

Gibbons have slow life histories, with similarities in flexible social and mating systems between the small and large apes (see Reichard & Barelli, 2008, for further discussion). Within gibbons, weaning is proposed to occur between 24-30 months, with puberty beginning around five years of age and sub-adults dispersing from their natal group around ten years of age (Brockelman, Reichard, Treesucon & Raemaekers, 1998; Burns & Judge, 2016; Reichard et al., 2012). Average age at first reproduction for *H. lar* at a long term field research site (Khao Yai, Thailand) is around ten years with inter-birth intervals of ~3.4 years (Reichard et al., 2012). However, ecological and social environments may influence the timing of these life history milestones. Burns and Judge (2016) conducted a long term assessment of behaviour development in *H. moloch*, *N. leucogenys*, and *S. syndactylus* groups, finding plasticity in the development of sexual, physical and social maturity within these species. Physiological and behavioural maturation was found to be slower for individuals remaining in their family group, with Burns and Judge (2016) suggesting this may allow for longer exploitation of parental territory and support. In contrast, faster development was found in individuals who dispersed at a younger age, potentially allowing for increased reproductive success.

1.4 Physical cognition within the small apes

1.4.1 Early investigations

Yerkes and Yerkes (1929) report early investigations of gibbon intelligence by Boutan (1914), examining problem solving abilities in a female white-cheeked gibbon (*N. leucogenys*). Boutan (1914) presented the human-reared subject with a series of five wooden puzzle boxes containing a food reward. Two sides of the boxes were solid, and two consisted of wire netting, with the reward placed inside and therefore visible to the subject. A hinged door on top of the box allowed for reward retrieval, with the mechanism required to open this door varying across conditions. Four boxes with visible opening mechanisms were presented, followed by one with hidden opening mechanisms. The manipulations required to open the box incrementally increased in difficulty. The first box required the subject to simply pull a wooden handle to lift the top door; the second introduced a wooden bar situated across the door that the subject had to pivot away before it could be opened. A third and fourth box introduced two and three pivoting bars respectively that had to be moved before the reward could be accessed through the door. The apparatus with a hidden opening mechanism consisted of a similar box, with a latch holding the door shut inside the box and attached to a wire which passed out of the box and through a tube, ending in a wooden handle. When pulled, this handle released the latch, opening the door and allowing access to the reward. The gibbon received no training and was placed alone in the testing room with the puzzle boxes whilst Boutan (1914) covertly monitored her responses.

Upon presentation of the first ‘visible’ task, the subject initially attempted to get to the reward through the side of the box before pulling the handle to open the top door and retrieving the reward. Her performance led Boutan (1914) to conclude the task was an extremely easy and simple one. When presented with a visible box containing one horizontal bar positioned across the door, Boutan (1914) reports the gibbon switched between attempting to lift up the door handle and the bar before apparently accidently knocking the swivelling bar, allowing the door to be pulled open. When subsequently presented with multiple horizontal bars holding the door closed, the gibbon slid the bars aside without attempting to lift them, apparently having mastered the task. Boutan (1914) reports the gibbon operated the bars quickly and without hesitation, suggesting understanding of the task demands. The gibbon also learned to pull the handle to open the door in the hidden apparatus. During an initial trial, the subject inspected various elements of the box, unsuccessfully attempted to pry open the door, and briefly touched the wire attached to the latch. A weight was attached to the wire during this trial, with Boutan (1914) considering the gibbon may have been successful in reward retrieval had the wire not been weighted. A second trial therefore removed this weight, and the gibbon successfully pulled the wooden handle and released the door. Boutan (1914) notes the gibbon recalled the successful behaviour required to open the various boxes over a period of three months, readily performing the correct sequences of movements to open the door following this break from testing.

Boutan (1914) proposed the gibbons’ performance was not indicative of trial and error learning, whereby repeated unsuccessful actions would occur before gradual learning of the correct response. Instead, Boutan (1914) suggested, the subject displayed immediate elimination of useless behaviours and sudden realisation of the task solution. However, several aspects of the experimental design make this claim that the gibbons’ behaviour was mediated by higher cognitive processes questionable. The discovery of the solution to the second visible mechanism box, with a pivoting bar holding the door shut, was noted by Boutan (1914) as accidental, rather than a deliberate reward directed action. It is possible the gibbon encoded the correct behavioural sequence from this chance occurrence and repeated the action across the successive experiments introducing a greater number of pivoting bars. Furthermore, it is noted that the nature of the hidden mechanism box necessitated trial and error learning. As the mechanism is hidden, the correct behavioural sequence for reward retrieval must first be discovered through chance manipulations. The speed with which the subject encoded the correct response is notable, reliably repeating the correct behaviour following success on the initial trial; however, trial-and-error learning cannot be excluded. Boutan’s (1914) assessments presented an interesting initial investigation into the problem solving abilities of this gibbon; however, as noted by Yerkes and Yerkes (1929), issues with the experimental procedure, a lack of task variety and incremental difficulty raises questions about the authors conclusions.

Yerkes and Yerkes (1929) also give a brief report of an early comparative assessment of intelligence conducted by Drescher and Trendelenburg (1927). The first of two tests described was a raking in task. These tasks present a rake shaped object and a food reward, both placed on a platform, with the food reward out of reach but in front of the rake end. Subjects are thus required to pull in the rake in order to slide the food reward to within reach. When this task was presented as a zero-order manipulation, whereby one object (the rake) is manipulated in order to produce an action on a second object (the reward placed in front of the rake) (Fragaszy, Visalberghi & Fedigan, 2004), the gibbon (sex and species unknown) succeeded. However, when required to re-orient the rake into the correct position behind the food reward before drawing it in, the gibbon failed to display the correct behaviour. Drescher and Trendelenburg (1927) report the gibbons performance was comparable to that displayed by monkeys on this task. In the second task, the gibbon was able to manipulate a puzzle box to retrieve a reward that required removal of the box cover in order to access the food reward placed inside. This subject capably removed the cover, continuing to do so when the box was repositioned, requiring the cover to be removed from an alternative direction. When the box cover was fastened with a bolt, the subject again succeeded in retrieving the reward, although it is noted this success only occurred after three days of testing. The gibbons' performance on this task was suggested to be inferior to the large ape species tested, and more comparable to that of monkeys. Drescher and Trendelenburg (1927) further note that despite interest in the task and food reward, the subject was shy and easily distracted.

There are few remaining assessments of cognitive abilities within the small apes prior to the 1960's. Harlow, Uehling and Maslow (1932, see also, Harlow, 1932) assessed one (*H. lar*) gibbon during a comparative investigation of 24 primates' ability to solve delayed reaction tasks, investigating initial learning capacity and the maximum delays at which subjects would still respond correctly. Delayed response tasks are often utilised to assess working memory capacity, requiring subjects to select a previously presented item from a range of options, following a variable delay period where the presented item is hidden (Rodriguez & Paule, 2009). Subjects included one orangutan (*Pongo pygmaeus*), one gibbon (*H. lar*), two capuchin monkeys (*Cebus capucinus*) one lemur (*Varecia variegata*), and a range of Old World monkeys. Old World monkeys included four guenons (*Cercopithecus* sp.), eight macaques (*Macaca* sp.), one mangabey (*Cercocebus* sp.), three baboons (*Papio* sp.) and three mandrills (*Mandrillus* sp.).

The delayed response task required subjects to choose between two opaque containers, one of which had previously been shown to contain a reward, either instantly following baiting or after a delay period (5, 15, 30, 60, 120, 180, or over 180 seconds). Subjects were permitted ten attempts at each delay length before they were considered as failing the task. A definitive criterion was not set for failure on the task, but determined by the experimenters as a marked increased in the number of errors and a negative emotional reaction to the task, such as striking at the apparatus or experimenter. The lemur and New World monkeys performed poorest, with the lemur achieving 57% success at a maximum delay of 30 seconds and the two capuchin monkeys failing at a delay of 15 seconds, achieving 65% and 73% correct choices. The majority of the Old World monkeys tolerated delays of up to 120 seconds, with performance at this delay ranging from 47-86% correct selections. The gibbon's performance was poorer than would be expected given their phylogenetic position. This subject achieved lower numbers of correct selections than several Old Wold monkeys, performing poorer than all adult baboons and six of the eight macaques. Percentage of correct selections for the gibbon subject remained above 80% for up to 30 second delays but fell to 71% correct after a 60 second delay, and 66% after 120 seconds. The orangutan subject did not complete testing; however, this individual succeeded during delays of 600 seconds and maintained a performance of 90% correct selections at a 120 second delay period. From these findings, Harlow et al. (1932) suggested a grouping of primate species based on their response capabilities, with lemurs and New World monkeys grouped together as 'lower primates', followed by Old World monkeys and gibbons, classified as 'intermediate primates', and finally large apes, considered to be 'higher primates'. However, Harlow et al. (1932) report the gibbon displayed low motivation levels throughout testing, being reluctant to respond to delays of a minute or more and displaying disinterest in the task. Thus, the gibbon's poor performance may have been a result of this indifference to the experimental testing, rather than reflecting an intellectual deficiency.

Individual differences in personality traits and temperament may affect both approach style and performance on cognitive tests. In particular, differences in exploration may relate to differences in learning, with 'fearful' or 'shy' individuals often omitted from testing procedures due to a reluctance to engage with the task, rather than a lack of ability (Carere & Locurto, 2011). Individuals with traits relating to 'openness' may be more curious and exploratory in response to novel situations, being more willing to engage in cognitive testing and, as a result, completing tasks faster than other subjects (Morton, Lee & Buchanan-Smith, 2013). For example, Coleman, Tully and McMillan (2005) found temperament correlated with rhesus macaque's (*Macaca mulatta*, n = 20) training success during a simple task. Subject's temperaments were first assessed by measuring responses to a novel food object, with individuals categorised in to three temperament levels; inhibited, moderate or exploratory. Subjects were subsequently trained to touch an object situated outside their enclosure in order to receive a food reward. Results showed those individuals categorised as inhibited took longer to produce the correct behavioural response. Likewise, Morton et al. (2013) found individual differences in personality traits of capuchin monkeys (*Sapajus apella*, n = 18) related to both task participation and performance. Five personality dimensions were measured, including assertiveness, openness, neuroticism, sociability and attentiveness. Two tasks were presented, both of which required subjects to choose between entering one of two testing compartments. During the first task, a food reward was placed, out of reach, in front of one of the compartments. Subjects were required to enter the compartment with the food reward directly in front of it in order to be handed the reward by the experimenter. In the second task, two opaque cups differing in size were placed in front of the compartments. A food reward was placed under the larger of these two cups, with subjects now required to enter the compartment with the large cup placed in front of it in order to be handed the reward.

Individual's scores on openness positively related to differences in task participation. Additionally, monkeys that consistently performed well on both tasks showed significantly higher openness scores and lower assertiveness scores compared to others. Findings such as these highlight the importance of distinguishing between a lack of willingness to engage with a task, potentially due to personality traits, and a lack of ability. As noted by Coleman et al. (2005), different testing methods may be needed in order to address specific individual needs. Berkson (1962) sought to investigate gibbons' motivational level, determining food preference in three gibbons (*H. lar*) before presenting a delayed response task. Food preference was scored by means of a simple choice task, with raisins and grapes being the most favoured foods (see also, Maslow, 1933). Berkson (1962) then assessed whether performance on a delayed response task would vary dependent on reward quality. Subjects were presented with a centrally placed food reward, which was subsequently moved to either the left or right of a table and covered with a container. A delay of 0, 5, 10, or 20 seconds was enforced before subjects were permitted to make a selection, with the food reward presented being either high, medium or low preference. An effort was made to control subjects' appetite, with their last feeding session occurring either 0, 2, 4, or 6 hours before testing.

Performance on the task decreased as delay period increased, with low quality rewards producing lower accuracy levels compared to medium or high-quality foods. Motivation levels throughout testing were high, with similar trial completion levels across the varied food deprivation conditions. Berkson (1962) suggests the differences in deprivation level were not sufficient, conducting a subsequent experiment to further investigate the effect of appetite on performance. The same task as previously used was presented, with length of food deprivation being moderated. Subjects showed an increased willingness to take part in testing after longer periods of food deprivation, with performance levels improving slightly as appetite increased. Berkson (1962) therefore concludes that the low motivational levels observed previously in gibbons may be counteracted by the regulation of appetite and use of high-quality rewards, highlighting the need for testing procedures to take in to account species-specific reactions to testing situations.

1.4.2 Object manipulation

In line with the suggestion that differences in personality traits and temperament may affect task performance; it is also of importance to ensure tasks are suited to species specific morphology. Gibbons' hands have been described as 'hook-like', with elongated curved fingers, a reduced thumb and a deep cleft between the thumb and first finger allowing for an increased grip span (Preuschoft, Schönwasser & Witzel, 2016; Tuttle, 1969). This specific morphology may be well suited to gibbon's arboreal environment and primary mode of locomotion, brachiation; however, the elongated digits may lead to difficulty in manipulating small objects (Prime & Ford, 2016). Apes use a variety of grips to pick up objects, with distinctions made between 'precision' and 'power' grips. During precision grips the thumb tip may oppose one or more finger tips whereas power grips involve all digits, including the thumb, being used to hold the object against the palm (Christel, 1993). In contrast to earlier reports suggesting gibbons display a smaller variety of grasping behaviours than the large apes (e.g., Christel, 1993), Prime and Ford (2016) found gibbons were capable of executing both precision and power grips at a comparable level to large apes. A variety of objects varying in size, shape, colour and texture were presented in the enclosures of two pairs of gibbons (*N. leucogenys*, n = 2, *H. lar*, n = 2). This included structural objects such as bars or branches, food containers, food objects, and both hard and soft non-food objects such as pet toys, clothing, and plush teddy bears.

Gibbons were found to use their hands in a variety of manners, capably manipulating objects regardless of the size or shape of the item. Differences were found in thumb use between the small and large apes, with gibbons using their thumbs to manipulate and probe at objects in a manner usually reserved for the index or middle finger in large apes (Christel, 1993). Gibbons also used their thumb to hold small food objects firmly to the palm whilst using their fingers to sustain their weight during suspension or locomotion, with Prime and Ford (2016) suggesting this ability may be an adaptation of these primarily suspensory apes to their arboreal environments. Thus, differences in hand anatomy may result in slightly different manipulatory techniques between the small and large apes; however, it may be expected that gibbons would manipulate objects to the same degree as other apes, if given suitable apparatus.

Early work assessing the capacity for object manipulation across primates found gibbons displayed more exploratory behaviours than lemurs, New World or Old World monkeys, but had a smaller behavioural diversity than large apes. Responses to novel objects were assessed in 11 rhesus macaques (*M. mulatta*) and 11 gibbons (thought to be eight *H. lar* and three *H. pileatus*) by Bernstein, Schusterman and Sharpe (1963). Subjects were placed into a restraining cage and presented with a range of objects, including an unfamiliar conspecific, a white rat and a range of toys. There were marked differences in responses between the gibbons and monkeys.

Although most individuals displayed avoidance of some of the unfamiliar toys, ten of the gibbons gently touched the objects, remaining active in their presence. In contrast, the monkeys displayed threatening behaviour, with the only contact being to bite or pull at the objects, whilst some individuals remained motionless. Most subjects actively avoided the rat, although a greater number of gibbons made some contact with the animal, and one gibbon held the rat between its legs in what Bernstein et al. (1963) describe as a 'maternal manner'. Ten gibbons touched the unfamiliar conspecific, compared to only five monkeys, who displayed behaviours such as fear grimacing and screeching. When an unfamiliar human stood outside subjects' home cages, the monkeys either hid from or exhibited threat behaviour towards the individual, with only two individuals accepting food offered. Seven gibbons attempted to touch or interact with the human, with all individuals remaining within view and accepting food. Gibbons were also found to be more active when placed in an unfamiliar enclosure and to consume food in a more leisurely manner compared to the hurried feeding of the monkeys. Bernstein et al. (1963) conclude the monkeys' response repertoire centred around submissive or dominance asserting behaviours, with the unfamiliar situations eliciting stereotypical emotional responses. In contrast, the gibbons performed more exploratory behaviour, actively attempted to escape when agitated and displayed a greater depth of emotionality overall. Bernstein et al. (1963) considered that gibbons may be less influenced by hierarchical relationships than macaques, potentially leading to the fewer dominate or subordinate behaviours displayed by these individuals.

A more detailed assessment of behavioural diversity was conducted by Parker (1974), investigating the range of behaviours displayed in response to a simple manipulable object in the form of a nylon rope. Ten species of primate were assessed, including two species of lemur (*Lemur catta,* n = 2, *Eulemur macaco*, n = 2), five monkey species (*Ateles geoffroyi*, n = 4, *Cebus* sp., n = 4, *Trachypithecus cristatus*, n = 4, *Macaca nemestrina*, n = 4, *Cercopithecus mona*, n = 4), one species of gibbon (*H. lar*, n = 4), and three species of large ape (*Pan troglodytes*, n = 4, *Gorilla gorilla*, n = 4, *P. pygmaeus*, n = 4). Two dimensions of manipulative behaviour were analysed, including the body part used and the action performed. Body part was used was grouped into 111 categories, including categories such as head and neck, hands and fingers, and arms. The action performed was grouped into 8 main behaviours, including; picking up/holding, pulling, pressing, throwing, mouthing/striking, shaking/waving/rubbing, forming/wadding/twisting/wrapping, draping and miscellaneous, with these behaviours further dived into 173 categories. A new response was recorded each time there was a change in one or more of these categories, for example, a new action or new body part used to manipulate the object.

The three large ape species displayed the highest mean number of actions and manipulated the rope with the highest diversity of body parts. Gibbons displayed the fourth highest means across these measures, followed by the lemurs in body part diversity, and macaques in number of actions. Parker (1974) further assessed the number of ways in which different body parts were combined with different actions in behavioural combinations. Large apes again displayed the greatest behavioural diversity, with a higher number of body part and action combinations found in these species. Chimpanzees displayed fewer combinations than gorillas or orangutans; however, chimpanzees performed nearly three times as many behavioural combinations as gibbons. The number of combinations displayed by gibbons was closer to that found in macaque subjects, both of which performed better than the remaining lemur and monkey species. Parker (1974) proposes the larger relative brain size of the apes could lead to greater behavioural diversity, with this diverse repertoire of available behaviours allowing for increased learning and problem solving abilities in these species.

Given the increased exploratory and behavioural diversity displayed within apes, it is important to provide captive populations with problem solving tasks, both providing enrichment and encouraging natural foraging behaviours. Cheyne (2009) assessed gibbons (*H. muelleri*, n = 4, *H. agilis*, n = 2, *N. gabriellae*, n = 4, *H. lar*, n = 2) responses and success at reward retrieval when introduced to a novel maze-like food puzzle, requiring object manipulation in order to retrieve a food reward. Gibbons were housed and tested in pairs, with the puzzle feeder attached to the outside of the enclosure mesh within the indoor section of each pair's enclosure. The puzzle consisted of three clear tubes attached on top of one another, with holes cut out along the length large enough for subjects to insert their finger and manipulate a food reward. The puzzle was baited at one end of the top tube, requiring subjects to manoeuvre the reward along this tube and then the middle tube before retrieving the food from a larger hole placed in the far end of the bottom tube. The apparatus was initially presented un-baited for half a day, allowing subjects to acclimatise to the presence of the puzzle feeder. Subjects were then given 24 hours access to the puzzle once baited. Two main observation periods were conducted, one in the morning, directly after presentation of the reward, and the second in the afternoon. Reward location was also noted the following morning to assess for any movement throughout the night.

All subjects investigated the puzzle, although only 6/12 successfully retrieved one or more rewards. Time taken to approach the puzzle varied from instantly to over five hours, with Cheyne (2009) suggesting some gibbons may be inherently wary of novel objects and should be given an acclimation period before being expected to interact with such items. Although most subjects used either a thumb or index finger to manipulate the reward, two of the gibbons, a sub-adult and a juvenile, attempted to use external objects. They successfully manipulated the reward with these objects, with one gibbon using a twig that they had broken from a branch and the other attempting to use a runner bean. Despite this success, neither made continued use of these objects, with Cheyne (2009) noting the dexterity of a finger was more suited to the task requirements. Subjects' age was suggested to play a role in the level of interaction with the apparatus, with younger individuals spending longer interacting with the task. Cheyne (2009) notes the greater rate of cognitive development present in younger gibbons as they learn about their environment perhaps makes these individuals more likely to attend to novel objects.

As noted, it is important to ensure tasks are suited to species-specific anatomy when conducting cross-species comparisons of cognitive abilities. Beck (1967) considers that the low motivational and performance levels often reported in gibbons may be due to the tasks used not being particularly suited to their morphology. Modifying procedures developed by Köhler (1925), Beck (1967) presented a range of string-pulling tasks to assess problem solving abilities in four gibbons (*H. lar*, n = 3, *H. pileatus*, n = 1). The apparatus consisted of a board positioned outside subject's enclosures, with lengths of string rising from the board and passing into the enclosure at approximately shoulder height. The elevated height of the string from the table made this task more suited to gibbon anatomy. Three problems were presented, the first consisting of three conditions. Presentation order was varied across subjects, in an attempt to distribute the effects of practice and fatigue.

Condition 1a presented a direct string pulling problem, with a piece of food attached to the end of the string that subjects were required to pull in to retrieve the reward. Condition 1b was identical to that of 1a, except the food was not attached to the string, with subjects therefore unable to retrieve the reward. The third condition, 1c, presented the same set up as 1a, with the addition of a larger food reward placed beyond subjects reach. Subjects could retrieve the smaller reward only, by pulling in the string. All subjects correctly pulled in the string to retrieve the reward during 1a, continuing to do so during

1b, when the reward was unattainable. Although this performance suggests subjects may have failed to understand the relationship between the string and food, Beck (1967) proposes the gibbons were capable of discriminating between the two conditions based on behavioural differences across tasks. During task 1a, subjects monitored the rewards movement whilst slowly pulling in the string, whereas in 1b they quickly pulled the string in and often spent time manipulating it. Beck (1967) suggests subjects pulled in the string in 1b due to the incentive value of the string itself. Time to retrieve the reward was longer in 1c than 1a, suggesting the additional reward may have distracted subjects. However, few instances of reaching for the unattainable larger reward occurred, with Beck (1967) suggesting subjects were aware of their physical limitations and that the reward was out of reach.

The second problem introduced a peg fixed to the surface of the testing board, which remained positioned outside subject's enclosures. One end of the string was attached to this peg, with the free end rising from this central peg and extending into the enclosure to either the left or right of this point. A reward was attached along the strings' length at a position out of the subjects' reach. In contrast to the earlier condition, subjects could not retrieve the reward by simply pulling in the string, but were required to move the string perpendicular to the enclosure, causing the food reward to travel in an arc to a point where it was within reach. Four of the subjects displayed proficiency at this task, with Beck (1967) suggesting performance was similar to that obtained by Köhler (1925) whilst assessing chimpanzees.

Within the third problem, the string was presented tied to the enclosure, passing down and around the peg secured to the board, before the loose end returned back into the enclosure at the same height as the tied end. A reward was secured to the length of the loose end, out of subjects' reach. In order to retrieve the reward, subjects had to pull the tied end of string, initially moving the food away from themselves, before passing around the peg and then coming into reach. Time to completion was longest on this problem, suggesting the task was challenging for the gibbons. It may be the requirement to initially move food away from them caused difficulty, as this is something they are unlikely to encounter in their natural habitat. On 11 of the 12 completed trials, subjects attempted to pull the incorrect loose end first, spending a relatively long time doing so before correctly pulling the tied end. On 8/12 trials, subjects spent a period of time engaged in non-task related behaviours following several initial unsuccessful attempts at reward retrieval. Upon returning to the task following this break, the first behaviour displayed was the correct response. Beck (1967) therefore suggests that subjects solved the problem insightfully, with the problem solution appearing suddenly following a period of quiet contemplation. However, subjects' performance could be the result of trial and error learning, with the repeated unsuccessful manipulations leading to learning the correct sequence of behaviours.

Further assessment of object manipulation skills within gibbons using a task suited to their morphology has presented variations of the raking in task briefly reported by Drescher and Trendelenburg (1927, as cited in Yerkes & Yerkes, 1929). Cunningham, Anderson and Mootnick (2006) assessed the number of unproductive actions occurring before reward retrieval and latency to solution during a raking in task, assessing if gibbons displayed an insightful understanding of the task, or, as in Beck's (1967) string problems, required trial and error learning. The task presented a rake shaped object, consisting of a rod with a rectangle attached to the end positioned on a table outside subjects' enclosures, with the rake handle extending into the enclosure. In line with Beck's (1967) suggestions, this rake positioning allowed for easier grasping of the rake handle, given gibbons' elongated hands. A food reward was placed directly in front of the rake end, with subjects able to pull in the rake and slide the reward into reach. Four hoolock gibbons (*H. leuconedys*) completed testing, with all subjects successfully retrieving the reward during the first presented trial. Time taken to retrieve the reward generally decreased across trials, as did the number of unproductive manipulations, which occurred infrequently throughout testing. One subject made only one ineffective manipulation across testing, doing so in the first trial. The lack of repeated unproductive actions and rapid success suggests learning was not achieved through trial-and-error, with Cunningham et al. (2006) proposing subjects may be capable of mentally representing the problem before acting.

Prior exposure to objects allows individuals to learn about object affordances, potentially increasing both the likelihood of object interaction and the efficiency of future manipulations. Cunningham, Anderson and Mootnick

(2011) presented a similar raking in task to Cunningham et al., (2006), assessing the effect of prior exposure to the experimental apparatus on gibbon's abilities during this zero-order task. The task was as described above; a rake was placed on a testing platform with the handle extending into the enclosure, and during test trials this rake was baited with a food reward. Subjects included 22 gibbons (*H. leuconedys*, n = 4, *N. leucogenys*, n = 3, *S. syndactylus*, n = 3, *H. moloch*, n = 4, *H. pileatus*, n = 5, *H. agilis*, n = 3) split into two groups, either prior exposure (PE), or no prior exposure (NE). For 7 days preceding testing the apparatus was placed outside the enclosures of the PE group, as it would be presented during testing. Subjects were free to manipulate the rake during this time; however, the apparatus were not baited. The NE group were exposed to the apparatus for the first time at the start of testing. Of the 22 gibbons assessed, 14 successfully retrieved the reward, with no difference in the number of successful individuals between PE and NE groups. Subject's age also had little effect on performance, with the mean age of successful subjects being similar to that of unsuccessful individuals. Despite a slightly lower time to solution in the *Hoolock* subjects, there was no significant difference between genera. Time taken to retrieve the reward, and time to first physical contact with the rake, did not vary significantly between successful individuals in the two groups. Subject's willingness to investigate novel objects did not appear to be affected by prior experience, or a period of acclimation, as suggested by Cheyne (2009).

When split by sex, there was no significant difference in latency to first physical contact, although the males in the PE group generally took longer to touch the rake than those in NE. Cunningham et al. (2011) suggest habituation to the non-rewarded apparatus presented during the prior exposure may have decreased interest in the male PE subjects, resulting in an increased time to first contact. Time taken to solution differed between the PE and NE groups for females but not for males. Females in the NE group took longer to retrieve the reward than both females in the PE group and males in both groups. It may be that differences in risk-aversion between the sexes led to the observed differences in time to solution. Cunningham et al. (2011) note that females may be more cautious in manipulations of novel objects, and as such may have benefitted more from the prior exposure where the neutrality of the objects can be learned, whereas object experience did not improve task efficiency in males.

Gibbon's appear capable of understanding the directly perceivable relationship between the rake and food reward during these zero-order manipulation tasks and the string pulling tasks presented by Beck (1967). However, the unsuccessful manipulations occurring before task solution suggests task relevant experience may be required in order to learn about these relationships. A brief report of a study by Inoue and Inoue (2002) also found a young gibbon (*H. lar*) succeeded during a raking in task, reliably selecting a baited rake. This task presented two rakes, placed in two separate 'lanes' on the testing platform. Both rakes were functional for reward retrieval; however, only one rake was baited. Inoue and Inoue (2002) further tested this gibbons understanding of the causal relationships between the rakes and food rewards by introducing a trapping hole, cut into the surface of one of the lanes (see Povinelli & Reaux, 2000, for further discussion of the trap-table task). If the rake placed behind this trap was selected, the reward would fall into the trap and be unobtainable. Successful performance on this task requires subjects to anticipate the interaction between three factors; the rake, food reward, and the surface of the table. Inoue and Inoue (2002) found the gibbon successfully selected the rake placed on the side of the apparatus without a trap. The gibbon continued to select the correct rake when both sides of the apparatus contained a trap, one placed in front of the rake and reward, and one (ineffective) placed behind, suggesting the subject recognised the spatial relationships between the rakes, rewards and traps.

Following the initial assessment described above, Cunningham et al. (2006) presented a more detailed assessment of gibbons' abilities during a trap table task. A training phase presented subjects with a choice task, requiring them to select the baited rake from a choice of two or three. Apparatus were similar to that described above, with either two or three rakes now being presented adjacently, one of which was baited. Subjects displayed high levels of accuracy on this task, selecting the baited rake on 156/160 trials. There was no significant difference in time taken to respond between the two and three choice conditions. Within the test phase, two sets of apparatus were presented adjacent to one another, one of which now included a trap. The trap consisted of a hole cut into the table surface, with the reward and rake end presented behind the trap, rendering reward retrieval impossible. Presentation side of the trap was randomised, with subjects required to select the non-trap side to retrieve the reward. Two of the four subjects performed significantly above chance, with one correctly selecting the non-trap side on 100% of trials and the other 74%. The remaining two subjects selected correctly on 64% and 56% of trials. The successful gibbons performed comparably to chimpanzees, which averaged 99% correct selections over 20 trials, during a similar task presented by Povinelli & Reaux (2000).

Despite successful performance by two subjects, Cunningham et al. (2006) note that, in contrast to Inoue and Inoue's (2002) suggestion, it is unclear whether behaviour was directed by comprehension of three-way causality, or was the result of learned associations between, for example, the unobstructed table and reward retrieval. Performance generally improved throughout testing, suggesting subjects may have learnt to avoid the trap over time. As with gibbon's performance on raking in tasks not incorporating a trap, repeated task experience may be needed in order to learn about the interaction between the rake, reward and trapping hole.

1.4.3 Tool use

True tool use is currently defined as '*the external employment of an unattached or manipulable attached environmental object to alter more efficiently the form, position, or condition of another object, another organism, or the user itself, when the user holds and directly manipulates the tool during or prior to use and is responsible for the proper and effective orientation of the tool'* (Shumaker, Walkup & Beck, 2011, p. 5). This updated the previous definition of tool use (see Beck, 1980) by now considering both attached and unattached objects as possible tools. It is the quality of the manipulation of the object which is essential to distinguish between attached objects that can and cannot be used as tools. Attached objects which are manipulated in such a way to meet the other necessary criteria for tool use are defined as tools. For example, Shumaker et al. (2011) would not consider an attached branch used as an anvil during nut cracking to be a tool, since it was not directly manipulated. In contrast, an ape using an attached branch in order to pull to within reach a second branch would be considered tool use, as the remaining criteria are met.

The phrase ‘to alter’’ implies goal-directedness, with Shumaker et al. (2011) noting few would disagree that tools are used purposively to achieve a goal. Some (e.g., St Amant & Horton, 2008) regard manipulations in which the form, position or condition of something is not altered, despite an individual’s attempt to do so, as unsuccessful tool use (for example, a failed attempted at cracking open a nut). However, Shumaker et al. (2011) consider these behaviours to be tool use, highlighting the ontogeny of tool using behaviour whereby juveniles may learn to use tools through repeated unsuccessful attempts. Although the current definition incorporates this cognitive construct of purposiveness, Shumaker et al. (2011) note that there is no obligatory relationship between tool use and causal understanding. Thus, tool use need not evidence the user’s causal understanding of the relationships or mechanical dynamics of the objects involved. Gibbons have been proposed to recognise causal relationships in the string-pulling experiments presented by Beck (1967). However, as subjects did not produce the proper orientation between the tool and goal objects, this would not be classified as tool use. Similarly, the raking in and trap table tasks previously discussed (Cunningham et al., 2006, 2011; Inoue & Inoue, 2002) do not represent an assessment of true tool use as subjects are not required to properly orient the rake before using it to draw in a food reward.

There are few existing reports of tool use in wild gibbons, consisting of breaking off and dropping branches, and shaking branches during vocalisations (see Shumaker et al., 2011, for a detailed overview). Geissmann (2009) reported a captive female gibbon (*H. lar*) slammed the door of her sleeping compartment during the climax of her great call, suggesting this represented a case of tool use within the small apes. The individual moved to her sleeping compartment briefly before her great call phrases, half shut the sliding door, and then slammed this door open at the climax of her great call, exiting the sleeping compartment and performing a brief locomotor display. In line with previous definitions of tool use (see St Amant & Horton, 2008), Geissmann (2009) considered this behaviour to be tool use as the gibbon controlled a freely manipulable object to mediate the information perceived by potential competitors. However, given Shumaker et al.'s (2011) current definition, it is debatable whether this behaviour may be considered tool use given that the door was not used to act upon, or alter, another object or organism, but to produce a sound.

The few experimental assessments of tool use in gibbons have found little evidence for true tool using abilities in these apes. Drescher and Trendelenburg (1927, as cited in Yerkes & Yerkes, 1929) found a gibbon failed to position a rake in an effective orientation during a raking in task; however, few details are given of this assessment. Cunningham (2006) presents a more detailed assessment with this task, investigating the spontaneous understanding of task requirements in six gibbons (*H. leuconedys*, n = 2, *N. leucogenys*, n = 2, *H. moloch*, n = 2) with previous experience a zero-order raking in task. Experiment 1 required subjects to use a T shaped rake to retrieve a food reward placed out of reach. The rake was either positioned in the correct alignment for reward retrieval, as in a zero-order manipulation, or offset by approximately 5cm or 30cm to either the left or right of the reward, requiring re-orientation before use. Subjects reliably raked in the reward during the zero-order manipulation but failed to successfully re-orient the rake during offset trials. The best performing subject retrieved 55% of rewards when the rake was offset by 5cm, with only 15% of rewards retrieved across all subjects during the 30cm offset trials.

Subjects did display some manoeuvring of rakes during offset trials in Experiment 1; however, it was not clear whether these behaviours were intentionally directed towards reward retrieval. A different apparatus was used in Experiment 2, allowing for clearer assessment of behaviours. Five gibbons (*N. leucogenys*, n = 2, *H. pileatus*, n = 3) were presented with a task similar to Experiment 1; however, the food reward was now placed in a dish, with subjects required to manipulate a rod with an X shaped end, hooking this into the dish and pulling in the reward. During test trials, where the rake end was positioned at an angle of approximately 45° to the reward dish, only two subjects successfully retrieved any rewards, doing so by using the rod in a sweeping motion in order to push the dish to within reach, rather than by the expected 'hooking' motion.

Overall, gibbons have displayed little evidence of true tool use, failing to re-orient rakes for effective reward retrieval. This suggests gibbons do not spontaneously understand the relationship between a tool and food reward when this relationship is not directly perceivable, as it is during zero-order manipulations. It is possible that with further experimental experience, allowing for trial and error learning, gibbons may have learned the correct behavioural sequences necessary for reward retrieval. These findings contrast with reports of tool use by large apes, which have been found to use, and manufacture, tools to retrieve out of reach food sources (see Shumaker et al., 2011, for a detailed overview). Cunningham (2006) conducted a further assessment of tool using abilities in gibbons, assessing performance on a dipping task. This task requires similar behaviours to those displayed by chimpanzees during extractive foraging, where a stick-like object is used to retrieve food sources such as termites or ants from nests or other cavities (Shumaker et al., 2011).

Eleven gibbons (*H. leuconedys*, n = 3, *N. leucogenys*, n = 4, *S. syndactylus*, n = 1, *H. pileatus*, n = 3) were presented with a transparent box containing a liquid reward with four holes in the top allowing for insertion of dipping tools. Baseline observations occurred over three days, assessing gibbons' spontaneous development of dipping behaviour. The tools were placed in the apparatus at the beginning of each day. Following this, subjects were given demonstrations of the dipping task by the experimenter. No subject spontaneously developed dipping behaviour during the baseline assessments, with only one gibbon retrieving the reward following experimenter demonstrations. This success may not have been the result of directed behaviour, as the dipping stick slid into one of the holes in the box lid after being placed down by the subject.

The use of a transparent box allowed subjects to continually monitor the spatial relationship between the tool and food reward. However, subjects frequently removed the tools from the testing area, and as such, may have limited the time available for learning the necessary behaviours for reward retrieval whilst both tool and reward were perceptually present. Two further experiments presented modified apparatus, restricting the removal of the tools in an effort to facilitate learning. Experiment 2 found subjects (*N. leucogenys*, n = 2, *H. pileatus*, n = 2, *H. agilis*, n = 4) again failed to display dipping behaviour following a training period where the tools were not removable from the apparatus due to a disc attached to the end of the tool. These subjects also failed to re-insert the tools during Experiment 3, where the tools were removable from the apparatus but restricted to the testing area by means of a chain attaching them to the enclosure mesh close to the dipping box. Gibbons therefore failed to display tool use even with both the tool and goal object were simultaneously perceivable. Cunningham (2006) suggests the cognitive abilities of gibbons more closely resemble those of monkeys than large apes, with monkeys also failing to correctly orient tools during raking in tasks requiring true tool use (see Chapter 2 for further discussion of this research).

1.4.4 Summary

Gibbons have been found to display a greater behavioural diversity in response to novel objects than monkeys, with Parker (1974) suggesting this increased behavioural repertoire may result in improved problem solving abilities compared to lower primates. However, in contrast to their phylogenetic position, located between the monkeys and large apes, early investigations found gibbons performed comparably to monkeys during delayed response tasks (Harlow et al., 1932). Gibbons temperament during these tasks was frequently described as disinterested or reluctant to take part in testing (Drecher & Trendelenburg, 1927; Harlow et al., 1932), with performance levels potentially a reflection of motivation levels rather than a cognitive limitation within these subjects.

Research taking in to account gibbons' unique morphology, in particular their elongated hands, which makes picking up objects laid flat on a testing platform difficult, found gibbons capable of solving problems where the relationship between the manipulable object and a food reward are directly observable (Beck, 1967; Cunningham et al., 2006; 2011, Inoue & Inoue, 2002). However, once required to produce the correct relationship between an object and a food reward for success, gibbons perform poorly, suggesting true tool use may represent a limit in cognition within the small apes (Cunningham, 2006). Gibbons success with zero-order manipulations during raking in and trap table tasks (Cunningham et al., 2006; Inoue & Inoue, 2002) suggests an understanding of the relationship between the rake, food reward, and the substrate along which the reward moved. However, it remains unclear whether gibbons display an understanding of which properties of the rake are relevant for successful reward retrieval during this task. A review of understanding of the functionally relevant, and irrelevant, features of objects within primates will be discussed in Chapter 2, followed by an assessment of this ability in gibbons during Chapters 3 and 4.

Chapter 2: Attending to functionally relevant features of tools and planning for future tool use

2.1 Introduction

The use and manufacture of simple tools has been reported in a wide range of species from insects to mammals. In particular, numerous reports describe the use of tools to extract food sources by non-human primates, both in the wild and in captivity (see Shumaker, Walkup & Beck, 2011, for a full review). Fewer reports have explored how non-human primates represent tools, and more specifically their level of understanding of which properties are relevant when considering the functionality of a potential tool (see section 2.2). This chapter first presents an overview of research assessing the ability to attend to the functionally relevant and irrelevant features of tools in primates. Proposed frameworks of prospective cognition are then discussed (see section 2.3.2). Previous assessment of prospective cognition in primates has focussed on restrictive criteria and assessment of future thinking comparable to that displayed by humans (see Suddendorf & Corballis, 2007, 2010). The benefit of broader investigations, allowing for assessment of the degree of future thinking displayed across species, is discussed. Following from this, definitions of prospective behaviours are considered, differentiating between the range of different behaviours which do require a sense of the future, and those that do not. A review of research assessing planned future tool use across large apes and Old World monkeys is provided, before an overview of the aims of the current thesis.

Traditionally, non-human animal intelligence has been viewed as modular, or domain specific processes evolved to deal with specific problems within species natural habitats (Kaufman, Reynolds & Kaufman, 2019). However, positive correlations have been found among performance on diverse cognitive tasks within a variety of primate species, including humans, providing support for the presence of a 'general intelligence' (reviewed in Burkart et al., 2017). General intelligence may be broadly defined as the ability to show behavioural flexibility, think abstractly and comprehend complex ideas across domains (Damerius et al., 2019). Inter-specific differences in general intelligence have been reported. For example, Deaner, van Schaik and Johnson (2006) conducted a meta-analysis of primate cognition research, finding no genus outperformed others within specific paradigms, but that there was a difference in overall task performance. Large apes generally performed better than other lineages, with the small apes, Old World and New World monkeys displaying a trend for outperforming prosimians. General intelligence has also been found to correlate with self-control in both human (e.g., Meldrum, Petkovsek, Boutwell & Young, 2017), and non-human primates, with Beran and Hopkins (2018) suggesting increased inhibitory control may aid with cognitive processing and intelligent decision making. The across domain correlations of task performance in human and non-human primates supports the evolutionary continuity of domain general intelligence. Thus, although understanding of functionality and components of prospective cognition can be considered as solitary constructs, these abilities may be intricately linked, with assessment contributing to the broader understanding of cognitive abilities within the small apes.

2.2 Attending to the functionality of tools

Early investigations with New World monkeys suggested these monkeys were capable of distinguishing between functionally relevant and irrelevant features of cane shaped tools, rather than simply mapping individual objects to single functions in an associative manner (Hauser, 1997). The monkeys were presented with a training task requiring them to select and pull in one of two cane shaped rakes in order to retrieve a food reward. The canes were presented on either side of a testing platform, split into left and right sections, with a small food reward placed beside each cane. One cane required no re-orientation before pulling in, as the food reward was placed within the cane’s hook. This task was therefore a zero order manipulation, whereby one object is manipulated to produce an action on a second object (Fragaszy, Visalberghi & Fedigan, 2004). Zero order manipulations may therefore not qualify as true tool use, if the subject is not required to effectively orient the object used to manipulate the second object (see Shumaker et al., 2011). However, these tasks allow for closer investigation of what an individual understands about the interaction between two objects and have been widely used in tool-use research with animals. For the second cane, the reward was placed outside the cane's hook, requiring re-orientation to retrieve the reward, as in true tool use.

Two sets of experiments were conducted with cotton top tamarins (*Saguinus oedipus*), assessing an adult (n = 9) (Hauser, 1997) and infant group (n = 4, ages 4-8 months) (Hauser, Pearson & Seelig, 2002). There are few reported incidences of tool use in cotton top tamarins, or other *Saguinus* species (see Shumaker et al., 2011, for more details); therefore these assessments looked at understanding of functionality in a typically non-tool using species. Hauser and colleagues found both age groups preferentially selected the most efficient option, choosing the cane with the reward positioned inside the hook. However, on trials where the rake requiring re-orientation was selected, subjects displayed no attempt to orient the cane into the correct position for reward retrieval, suggesting a lack of understanding of the causal relation between the canes and rewards. Replication of this experiment with capuchin monkeys (*Cebus apella*) produced some differences in performance. Fujita, Kuroshima and Asai, (2003) found subjects (*C. apella*, n = 4) also preferentially selected the correctly oriented cane, however, performance appeared to be affected by the positioning of the two food rewards, with capuchins displaying a tendency to reach for the item closest to them. Cummins-Sebree and Fragaszy (2005) report subjects (*C. apella*, n = 6) also preferentially selected the correctly oriented cane as a group, however, successful reward retrieval was not restricted to selections of this cane. Subjects in this study occasionally retrieved the reward by manipulating the incorrectly oriented rake, using methods other than 'contained within the hook' to retrieve the reward. Capuchins therefore displayed more exploratory behaviour than tamarins, but both species appeared able to attend to the reward position, and the functional relevance of this position for reward retrieval. It remains unclear whether the monkeys understood the relational concept of the reward being positioned 'inside' the cane hook during this experiment, or were attending to directly observable factors (association between hook and reward). Experiments manipulating the perceptual features of the rakes were therefore presented to the monkeys, assessing whether subjects could attend to changes that affected functionality, ignoring irrelevant featural changes.

A second series of experiments presented both canes oriented so no manipulation was required before pulling-in the cane to retrieve the reward (Cummins-Sebree & Fragaszy, 2005; Hauser, 1997; Hauser et al., 2002). As such, all canes were useable for reward retrieval, however, the effort required to do so was manipulated by changes in perceptual features. A single change, in colour, shape, texture, or size was presented for each rake. For example, two canes were presented, one identical to the original cane except for a change in colour, and the second identical to the original other than a change in size. Colour changes did not affect rake functionality; however, the ease with which the cane could be manipulated was affected by changes in cane texture, and to a greater degree, changes in size. Changes in shape had the greatest effect on functionality. Both age groups of tamarins tolerated changes in the perceptual features of the canes, with all canes being selected at least once (Hauser, 1997; Hauser et al., 2002). However, tamarins preferentially selected canes where the perceptual changes least affected functionality. Canes with altered colour or texture were selected more frequently than those with novel shapes or sizes. Capuchins selected canes with colour changes over those with size changes, but did not display marked preferences for texture changes over shape changes (Cummins-Sebree & Fragaszy, 2005).

In Experiment 3, a new set of canes incorporating a combination of the previous featural changes were presented as in the initial experiment, with canes being positioned in such a way that re-orientation was required in order to retrieve the reward, or a simple pulling in motion would slide the food reward within reach. Overall, correctly oriented canes were preferred by all monkeys, with subjects tolerating the changes in perceptual features. On trials where two canes requiring re-orientation were presented, tamarins did preferentially select the cane that required the least manipulation in order to obtain the reward. However, there was no evidence of tamarins effectively manipulating these canes, with no successful reward retrievals in trials where two canes requiring re-orientation were presented. In contrast, Cummins-Sebree and Fragaszy

(2005) report capuchins frequently attempted to reposition incorrectly oriented rakes, occasionally succeeding at reward retrieval.

During the final experiment, familiar canes requiring re-orientation were presented with novel canes placed in positions that did not require re-orientation, assessing if subjects could attend to the functionality of rake configurations or would be guided be familiarity. All subjects again preferentially selected canes requiring no re-orientation before pulling in; suggesting novelty did not affect their ability to attend to the functionality of the presented rakes. As in Experiment 3, tamarins did not attempt to re-orient canes and failed to retrieve the reward on trials where a cane requiring re-orientation was selected (Hauser, 1997; Hauser et al., 2002). Capuchins again frequently attempted to re-orient canes and did so with increasing success when compared to Experiment 3 (Cummins-Sebree & Fragaszy, 2005).

Fujita et al. (2003) also found that capuchins continued to select correctly oriented canes when perceptual differences were manipulated. Following from the initial experiment described above, Fujita et al. (2003) conducted three further experiments presenting correctly and incorrectly oriented rakes. The second experiment again presented hooked canes, now of different colours, with subjects preferentially selecting the correctly oriented rake, showing no preference for either cane colour. The third and fourth experiments introduced pairs of canes in new shapes. By presenting canes of the same shape simultaneously but in different positions, Fujita et al. (2003) were able to more closely assess subjects understanding of spatial relationship between the cane and reward without the influence of other perceptual differences.

Canes were presented in either original or inverted positions, with the reward presented either ‘inside’ or ‘outside’ the cane. Thus, during Experiment 3, a cane presented in the inverted position was ineffective for reward retrieval, whereas canes presented in the original position were usable for reward retrieval only when the reward was positioned ‘inside’ the cane. Subjects performed significantly above chance, suggesting they attended to the spatial relationship between the cane and food reward. A trend towards selecting the cane with the reward situated ‘inside’, regardless of orientation, was found; however, this was not significant. A more substantial difference in cane shape was presented during Experiment 4. Both inverted and non-inverted canes were now usable for reward retrieval, if the reward was positioned 'inside' the cane. Furthermore, the distances of the two rewards from the subject were varied to further assess the difficulty in abstaining from selecting the closest object displayed by capuchins during Experiment 1 (described previously). Subjects again preferentially selected the correct cane, retrieving the reward on the majority of trials. Performance was poorest on those trials were the unattainable reward was positioned closer than the reward placed 'inside' of the cane, suggesting a bias towards selecting closer items limited capuchins abilities.

Overall, monkeys appeared able to attend to the functionally relevant features of canes, preferentially selecting canes positioned in effective orientations, and tolerating irrelevant changes to perceptual features. Furthermore, capuchins displayed the ability to attend to the spatial relationship between the canes and rewards. Attending to functionally relevant features did not appear dependent on prior experience, with the infant tamarins, who had minimal experience with object manipulation and no prior experience of experimental testing, performing comparably to the adult tamarins. Capuchins occasionally re-oriented the canes to retrieve the reward, and displayed more exploratory manipulations than tamarins, as would be expected given the greater number of reported tool using incidences by capuchin monkeys (*Cebus* sp.) (see Shumaker et al., 2011, for more details). These findings do not suggest the monkeys necessarily understood the causal relations between the canes and rewards, or why specific perceptual features were relevant to reward retrieval (see Penn & Povinelli, 2007, for a critical review of causal cognition). However, Hauser et al. (2002) proposed non-tool using primates may have an innate ability to distinguish between functionally relevant and irrelevant tool features, even in the absence of relevant experience.

This hypothesis was questioned by Spaulding and Hauser (2005), who suggested the repeated exposure to the canes, and training with these tools during previous experiments, may have provided sufficient experience for subjects to generalise to new tool sets. The innate ability to distinguish between tool features proposed by Hauser et al. (2002) may depend on prior experience in order to trigger recognition of functionally relevant features. To assess this proposal, Spaulding and Hauser (2005) presented tool choice tasks to tool naïve subjects in the absence of any training periods. Subjects included cotton top tamarins (*S. oedipus*, n = 5) and common marmosets (*Callithrix jacchus*, n = 6), for which there are few reported incidences of tool use (see Shumaker et al., 2011). A similar experimental set up was used as that presented by Hauser and colleagues (Hauser, 1997; Hauser et al., 2002). Subjects were required to choose one of two tools, baited with food rewards and placed on a testing platform, pulling in the tool to rake in a food reward. Similar tools to those used by Hauser and colleagues (Hauser, 1997; Hauser et al., 2002) were presented during Experiment 1, varying in shape, colour, size, texture, material and orientation (some tools required re-orientating before they could be used to retrieve the reward). Both species performed at chance levels during the first session of 20 trials, failing to distinguish between features of tools that affected functionality. Across the four sessions presented, the marmosets, but not the tamarins, began to significantly select the tool usable to obtain the reward, displaying evidence of learning and performing significantly better in the fourth session than the first.

A further experiment aimed to assess the effect of this experience when presented with a set of novel tools, varying in both shape and material. Tools were constructed of clay, rope, rubber, or woodchips, with the clay tools always being functional, the woodchip never functional, and the rope and rubber occasionally functional. Trials were now presented in blocks of 22, with four sessions given. Tamarins again failed to perform significantly above chance during the first session, but reached significance by the final session. Marmosets significantly selected the correct tool from the first session, and continued to do so throughout testing, generalising over a wider range of tool combinations than the tamarins who mostly succeeded on trials with the most distinct difference between tools (i.e., hard clay versus woodchip).

From these findings, Spaulding and Hauser (2005) propose both tamarins and marmosets do possess an innate mechanism for recognising functionally relevant features of tools; however, this mechanism requires experience for discriminations to be made successfully. Without experience, neither species displayed the ability to discriminate between features that affected functionality and those that did not. Throughout the experiments described above, subjects gained experience directly manipulating the canes. Santos, Miller and Hauser (2003) assessed whether cotton top tamarins (*S.*

oedipus, n = 9) and rhesus macaques (*Macaca mulatta*, n = 24) required this physical experience operating tools to learn about functionally relevant features, or whether they attended to these features without direct manipulation experience. An expectancy violation procedure was used, whereby subjects looking time was measured as they witnessed an experimenter using a tool. The perceptual features of this tool were manipulated across experiments, assessing if subjects displayed an increased looking time during unexpected outcomes; when a perceptually non-functional tool appeared to successfully move a food reward.

An L shaped tool was used by the experimenter to push a food reward along a platform until it fell down a ramp to a lower platform. Following habituation trials, where subjects could witness this action, they were given ‘new colour’ trials, where a tool with the same functionality but different colour was presented, and ‘new shape’ trials, using a non-functional stick-like tool. During these trials an opaque barrier blocked subjects’ view of the upper platform (where the tool interacted with the reward), with the experimenter surreptitiously sliding the food reward down to the lower platform in all trials, regardless of the tool used. This visual barrier allowed for assessment of the unexpected outcome of non-functional tools appearing as functional to subjects.

Both species looked longer during trials where a non-functional shaped tool was presented than when a functional tool of a new colour appeared to move the reward. The question of whether subjects were attending to tool shape, or simply looking longer at this change as it was more salient than colour changes was assessed during Experiment 2. Familiarisation trials with both tool types were presented before test trials, with the finding that both species (*S. oedipus*, n = 11, *M. mulatta*, n = 24) again looked longer at shape changes during test trials. More importantly, subjects spent the same amount of time looking at a new shape and a new coloured rake during familiarisation trials, suggesting the salience of the changes were not guiding behaviour. Experiment 3 removed shape differences between the rakes to further investigate subjects’ ability to attend to the functionality of the presented tools, rather than simply shape changes. Following familiarisation trials, the L shaped rake was either presented in the original orientation (functional) in a new colour, or inverted (non-functional) in the original colour. Subjects (*S. oedipus*, n = 12, *M. mulatta*, n = 30) tended to look longer at the new orientation than the new colour; however, this difference was not significant. Species differences were found, with tamarins looking longer during the inverted tool trials, but not macaques.

In line with the findings of Hauser and colleagues (Hauser, 1997; Hauser et al., 2002), tamarins therefore attended to featural changes that affected functionality, doing so without direct manipulation experience with the tools, whereas macaque subjects responded to shape, but not orientation, changes. Santos et al. (2003) consider that macaques may lack the capacity to understand functional relations between objects, which is surprising given the greater evidence for tool use reported in macaques (*Macaca* sp.) than tamarins (*Saguinus* sp.) (see Shumaker et al., 2011). It is possible that differences in the subject groups tested affected performance. The tamarin subjects had previous object manipulation experience, being housed in a university laboratory, and many of the same subjects participated in each experiment, whereas the macaque subjects were free-ranging, with different subjects participating in each experiment. Although both groups displayed a bias for attending to shape changes, as suggested by Spaulding and Hauser (2005), the experiential experience may have affected understanding of which features were functionally relevant for a particular tool. This proposal is further supported by findings from Santos, Pearson, Spaepen, Tsao and Hauser (2006) who conducted a series of tool using experiments with a group of cotton top tamarins (*S. oedipus*, n = 10), some of which had participated in object manipulation experiments, and experimentally naïve vervet monkeys (*Cercopithecus aethiops*, n = 5). There are few reported tool using instances for vervet monkeys; however these monkeys have been reported to use sticks and other objects to obtain out of reach items (see Shumaker et al., 2011). During an initial training task, which replicated that presented by Hauser (1997) described above, the vervet monkeys learned to select the correct cane faster than the tamarin group as a whole. However, the tamarins with previous experience learned to select the correct cane at the same rate as the vervet monkeys, doing so four times faster than the naïve tamarin group. Vervet monkeys' performance improved as testing progressed, again suggesting experience manipulating the rakes assisted in learning to attend to functionally relevant features.

Santos et al. (2006) assessed the monkeys' ability to take into account the three-dimensional relationship between the tools and the food reward, presenting a replication of an inverted tool task developed by Povinelli and Reaux (2000). This task was part of a series of tool using experiments investigating chimpanzees (*Pan troglodytes*) understanding of the causal relations between the properties of a rake and a goal object, which was a food reward (Povinelli, 2000, see Shumaker et al., 2011, for a review of chimpanzee tool use). As stated above, these tasks presented zero order manipulations, and are not defined as true tool use as the subject is not required to effectively orient the rake. The inverted rake task presented chimpanzees (*P. troglodytes*, n = 7) with two identical rakes, one orientated so it would be functional to draw in a food reward, and the other inverted so as to be non-functional. The rake ends consisted of two vertical prongs, attached to either end of a horizontal bar with the rake handle attached perpendicularly to the middle of this horizontal bar. Thus, when presented with the vertical prongs pointing upwards, the horizontal bar could be used to slide a food reward in to reach. When inverted, the reward would pass through the gap below the horizontal bar created by the rake resting on the two vertical prongs.

The chimpanzees' performance on this task was close to chance levels, suggesting subjects did not consider the functional relevance of the rake's orientation when making selections. Chimpanzees selected the functional rake on 57% of trials when first presented with the inverted rake task, with only 3/7 apes preferring the functional tool. Performance dropped to 50% correct selections during a replication of the task, with only 2/7 chimpanzees now preferring to select the functional tool. These chimpanzee results are consistent with Santos et al. (2006), who found tamarins, when presented with two identical tools, chose a functionally oriented rake over one positioned so as to be non-functional around chance levels (52% of trials). Vervet monkeys did learn to select the functional rake significantly above chance, suggesting they attended to the rake orientation; however, Santos et al. (2006) note performance levels were similar to chimpanzees' on initial presentation (60% correct selections).

Povinelli and Reaux (2000) considered the applicability of two potential models to account for the chimpanzees' performance during the inverted rake task. Firstly, the ‘perceptual containment model’ proposed subjects may have perceived the reward as being ‘contained in-front’ of the rake, without taking in to account the functionally relevant features of the individual rakes (i.e. inverted or not). Alternatively, the ‘physical contact model’ proposed the chimpanzees may have reasoned about whether the rake would come into contact with the reward, but required more experience to attend to the relevant features of the task. A series of experimental conditions were presented in a random order to assess the validity of these two accounts. The same rakes were presented as in the initial experiments, with the reward position, and rake orientation manipulated. Presenting the rakes in either a functional or non-functional inverted orientation assessed if subjects could attend to whether the rake would come in to contact with the reward or not. Positioning the reward either in front or to the side of the rake assessed if subjects would select the reward ‘contained in front’ of the rake, regardless of rake orientation.

Neither model fully accounted for subjects’ performance, with Povinelli and Reaux (2000) suggesting subjects may have attended to different aspects in the different conditions. The chimpanzees appeared to attend to the rakes orientation on some trials, while on others, they attended to whether the reward was situated in-front of the rake, regardless of whether it was inverted or not. On three of the presented conditions the two models predicted selections of different rakes. The first of these conditions replicated earlier tasks, presenting one functionally orientated and one non-functional (inverted) rake, both with the reward placed in front. The chimpanzees again did not select the functional rake significantly above chance levels. However, as would be predicted by the physical contact model, 5/7 subjects now preferentially selected this functional rake over the inverted rake, potentially as a result of repeated exposure to this test set-up. The second condition presented one functionally oriented rake with the reward positioned outside the rake, and an inverted rake with the reward position in front of the rake; as such pulling in either rake would not lead to reward retrieval. Subjects selected the functionally oriented rake slightly more often during this condition, in accordance with the physical contact. The third condition presented two inverted non-functional rakes, one with the reward positioned in front of the rake and one outside. Subjects now responded in line with the perceptual containment model, selecting the rake which was positioned behind the reward. Although the second and third conditions presented the chimpanzees with no correct option as neither of the rakes led to reward, subjects still made a selection on the majority of trials.

For the remaining three conditions, both models predicted selections of the same rake. Two conditions presented one correctly oriented rake with the reward positioned in front of it, paired with either a correct or incorrectly oriented rake with the reward positioned outside of it. Chimpanzees selected the correct rake across these two conditions, as would be predicted by both models. The final condition presented one correctly and one incorrectly oriented rake; however, the reward was now positioned outside both rakes, and therefore unattainable. Chimpanzees tended to select the correctly oriented rake in this condition, displaying some sensitivity to rake orientation. Thus, the chimpanzees may not have fully grasped the physics of functionality of the different rakes, although they were able to attend to some perceptual features of the rakes.

In summary, both tool using and typically non-tool using species have demonstrated the ability to discriminate between functionally relevant and irrelevant features of tools during two dimensional raking in tasks. However, as proposed by Spaulding and Hauser (2005), this may reflect a more general sensitivity, requiring task relevant experience in order to achieve reliable success. Tamarins (Hauser, 1997; Hauser et al., 2002), capuchins (Cummins-Sebree & Fragaszy, 2005; Fujita et al., 2003) and marmosets (Spaulding & Hauser, 2005) have displayed the ability to differentiate between rewards contained ‘inside’ tools, and those placed outside of the tool. Tamarins have also displayed sensitivity to tool orientation during an expectancy violation paradigm (Santos et al., 2003). However, when combining consideration of the position of the reward relative to the tool, and the orientation of the tool itself, in a three dimensional manner, performance is poorer across both typically non-tool using species (vervet monkeys & tamarins), and tool-using chimpanzees. Povinelli and Reaux (2000) suggest that the perceptual containment model, considering whether the reward is positioned in front of the tool or not, may account for behaviour of naïve subjects when first attempting problems such as the inverted rake task. As with the findings from two dimensional raking in tasks, repeated task experience may lead to greater consideration of the functional relevance of tool orientation.

2.3 Prospective cognition

Human's ability to escape the present, transcending the physical constraints of space and time to mentally travel, allows us to not only relive past events, but also to plan for future eventualities (Miloyan & Suddendorf, 2015; Tulving, 1985, 2005). Whilst episodic memory allows an individual to remember one's previous experiences, episodic foresight is proposed to confer the ability to mentally construct potential future experiences (Suddendorf & Corballis, 2007; Tulving, 2005). We do not remember the past with complete accuracy, indeed memory is often subjective, malleable and constructive (Clayton & Wilkins, 2017; Suddendorf, 2010). Memory's adaptive advantage may therefore lie in informing and supporting current, or future behaviour, rather than as a repository of the past (Corballis 2014; Suddendorf, 2010). As Suddendorf (2017) notes, the future has priority in terms of evolutionary fitness. Prospection is a flexible, generative system, often involving extrapolations from memory and semantic knowledge, allowing humans to combine and recombine these elements in ways that confer adaptive advantages (Suddendorf, Addis & Corballis, 2009; Suddendorf & Busby, 2003a).

Prospection may draw on various cognitive processes such as theory of mind, self-awareness and temporal order, thought to have developmental milestones around the age of 3-4 years in humans (Suddendorf & Busby, 2003a, see Suddendorf & Corballis, 2007, for a review of associated components). Developments relating to memory may precede those relating to foresight, with Suddendorf (2010) suggesting future simulations are less constrained than those relating to the past, requiring a different measure of creative construction. Although children may begin to display an awareness of the future as young as 2 years of age, episodic foresight is thought to emerge around 3-5 years, with these capacities continuing to develop throughout childhood and into adolescence (Atance & O'Neill, 2005; Suddendorf, 2017). Suddendorf and Corballis (1997, see also, Tulving, 2005) suggest the ability to mentally travel in time, both back and forth through a subjective time continuum, is a uniquely human trait. The relatively late emergence and early deterioration of mental time travel across the human lifespan supports its recent evolution, and human uniqueness (Tulving, 2005).

2.3.1 Planning in natural settings

If animals possess foresight and are able to consider future consequences of current actions, they may formulate plans, acting in the present moment to secure future benefits. Humans display flexibility in future-oriented behaviour, being capable of planning across diverse situations. The assessment of non-human animals planning abilities across different domains lessens the argument for innate behavioural predispositions guiding behaviour (Suddendorf & Corballis, 2010). Primates are adept at mapping and monitoring their home ranges, orienting travel paths towards valuable resources that are often out of view (Beran et al., 2016a). This detailed knowledge of potential food sources throughout a territory may improve foraging efficiency, allowing for planned, directed travel paths as opposed to random searches (Brockelman, 2009).

van Schaik, Damerius and Isler (2013) found the direction of Sumatran orangutans' (*Pongo abelii*, n = 15) long calls predicted subsequent travel direction, with new calls indicating subsequent changes in direction. Male orangutans were observed from their sleeping nest site in the morning until evening nest site, for a maximum of ten days. Of particular interest, the last long call given in a day was found to be indicative of travel direction the following morning, up to 22 hours after the call was given. Furthermore, neighbouring female orangutans adjusted their morning travel directions in relation to the males late night call (prior to any morning great calls). This communication of future travel plans is suggested by van Schaik et al. (2013) to demonstrate orangutans' ability to plan for the future in natural contexts. Observing female chimpanzees (*P. troglodytes*, n = 5) over 275 days, Janmaat, Polansky, Ban and Boesch (2014) found individuals situated their sleeping sites in the direction of food sources, and departed earlier in the morning when feeding on highly ephemeral figs. The duration of time between choosing a nest location in the evening and departure the following morning, along with the number of other activities conducted during this time (including sleeping) led Janmaat et al. (2014) to suggest this behaviour evidences future planning.

Chimpanzees did not choose sleeping trees closer to ephemeral trees than other feeding sites, potentially due to preference of particular tree species as sleeping sites, predator avoidance, or sleep disturbance by nocturnal foragers. However, selection of sleeping sites en route to ephemeral fruit patches, in combination with an increased approach speed to these trees, suggest these chimpanzees used memory for the location of particular trees to guide sleeping site selection. Gibbons may also possess a spatial cognitive map of their home range, using similar pathways within their territory repeatedly in order to make more efficient movements between preferred food sources. Asensio, Brockelman, Malaivijitnond and Reichard (2011) monitored the movements of 11 habituated gibbon groups (*Hylobates lar*) over five consecutive days per group, in order to assess whether daily travel paths were goal oriented. Data was collected on the location of food sources visited, intergroup encounters, duets, and those locations where the group’s travel direction significantly changed. Group home ranges were calculated, with all available sources of preferred food mapped. Assessment of which locations were, and were not visited, allowed a comparison between the groups actual travel path and the optimal route given the food sources spatial distribution.

Changes in direction of travel were most frequently associated with preferred food sources, with a small number of change-points occurring following inter-group encounters, and none coinciding with duet locations. Asensio et al. (2011) suggest gibbons’ daily travel paths are therefore directed by foraging, rather than social behaviours, with subjects’ next location being planned before leaving a current food source. This is supported by the large distance often found between change points, with these food sources generally being out of visible range. Gibbons’ relatively small and highly mobile social grouping may allow them to exploit their knowledge of their territory for maximum benefit (Brockelman, 2009). It is noted however, that these findings support the planning of movement to only one subsequent location, with further research needed to address whether gibbons are capable of planning multiple movements in advance. Assessment of route-planning allows investigation of future-directed behaviour in a natural context; however, these studies are limited by the inability to control factors affecting behaviour, both during observations and through a lack of knowledge of previous life histories (Thom & Clayton, 2016).

2.3.2 A framework of prospective cognition

Experimental research assessing prospective cognition within primates has frequently focused on the dichotomy of abilities between humans and non-human primates, applying strict criteria in order to assess if episodic prospection, or the human ability to mentally travel in time, is present across other species. Suddendorf and Corballis (2007, 2010) set out four main criteria for assessing mental time travel in animals. Firstly, the use of **single trials** avoids learning of stimulus-reward relationships. They further note observed behaviour should not be based on generalisations, but on the memory for one specific event. Secondly, **novel problems** should be used to minimise potential innate responses and the effect of previous learning histories. A distinct **temporal-spatial separation** between action and consequence should be implemented to avoiding cueing behaviour and to ensure behaviour reflects long-term memory. Finally, tasks incorporating a **range of domains** should be employed, reflecting the flexibility of human mental time travel. In addition to these criteria, the ‘Bischof-Köhler’ hypothesis has been proposed (see Suddendorf & Corballis, 1997), stating animals are unable to dissociate from their current mental state to envision a past or future one; in essence they are unable to anticipate future needs as distinct from those currently experienced. Suddendorf and Corballis (1997) suggest even flexible future oriented behaviour, such as the manufacture of tools for future use, may not be indicative of mental time travel as these behaviours may be tied to a current desire or motivational state.

Martin-Ordas (2016) critiqued these criteria, firstly addressing the use of single trials, questioning whether this criterion actually distinguishes between learning accounts and episodic events. The use of single trials may not necessarily rule-out associative learning; conversely, repeated exposure may be required in order to retain stimulus relevant information (Morris, 2001). This criterion also fails to account for human memories that often refer to general repeated events rather than unique events. Martin-Ordas (2016) also suggest the contextual detachment proposed by temporal-spatial separation does not provide a realistic assessment of episodic foresight. Some aspect of the event, be it perceptual stimuli or mental states, must cue the future oriented behaviour (Osvath & Martin-Ordas, 2014). Finally, although intuitive, the flexibility criterion leads to questions of what flexibility actually involves. Despite aiming to minimise innate or associatively learned responses, an individual may still combine knowledge of previous experiences to guide behaviour.

Thom and Clayton (2016, see also, Raby & Clayton, 2009) further criticise the Bischof-Köhler hypothesis, proposing it is neither a sufficient nor necessary indicator of prospective mental time travel; episodic prospection is not required in order to act for future needs, nor is dissociation from current needs required in order to imagine potential future events. Thus, Suddendorf and Corballis's (1997, 2007, 2010) criteria may lack clear definitions of what qualifies as episodic prospection in non-human primates. Indeed, much future planning conducted by humans fails to meet these criteria set for prospective cognition in animals (Paxton & Hampton, 2009; Raby & Clayton, 2009). Focus on these restrictive criteria may obscure assessment of the range of prospective cognition found across species. Many animals display behaviours that confer future benefits without requiring mentally pre-visioning or planning for a future event. Raby and Clayton (2009) proposed a framework of future oriented behaviour, categorising behaviours into those that require no sense of the future versus those that depend on prospective thinking.

Fixed action patterns, such as hibernation or migration suggest anticipation of future needs: however, these behaviours may be cued by environmental changes and depend on evolved instinctual mechanisms (Roberts, 2007; Roberts & Feeney, 2009a). These behaviours are inflexible, often being tied to a specific domain, and may be displayed before an animal has experienced the situation requiring the behaviour (Suddendorf & Busby, 2005). Likewise, learnt associations, such as conditioned responses or time-place associations, may have future consequences but do not necessarily require a sense of the future. Raby and Clayton (2009) note these behaviours may have a degree of flexibility, with the strength of the association being regulated by the reliability of a cue predicting a certain outcome. Learnt associations may still be future-oriented in that individuals may guide future events by adjusting current behaviour, with this goal directed behaviour considered a form of prospection (Suddendorf, Bulley & Miloyan, 2018).

Different species may possess different component processes of prospective cognition conferring fitness advantages specific to those species' ecological demands, and it is of importance to assess to what degree these variations allow for future planning (Raby & Clayton 2009; Thom & Clayton, 2016). Focus on restrictive criteria may overlook these subtle differences in prospective cognition found between species. Adopting the framework proposed by Raby and Clayton (2009) allows for a broader focus, investigating component processes and species-specific adaptations. Investigating the levels of prospective cognition displayed by different species will allow for comparisons across phylogeny, assessing the evolutionary emergence of these capacities.

Three categories of behaviour are proposed by Raby and Clayton (2009) to require a sense of the future; prospective memory, semantic future thinking, and episodic future thinking (see Szpunar, Spreng & Schacter, 2016, for an alternative framework of prospection). Prospective memory involves the formation of an intention that cannot currently be fulfilled, retention of this intention over a delay period, and then execution of the intended response at the appropriate future time (Evans, Perdue & Beran, 2014). Put more simply, prospective memory tasks require the necessary cognitive control to encode, retain and then implement intended future behaviours (Beran et al., 2016a). Assessments with Panzee, a language trained chimpanzee (*P. troglodytes*), have provided support for prospective memory in the large apes. Panzee succeeded in a task requiring her to remember to return a lexigram token representing a hidden food item (Beran, Perdue, Bramlett, Menzel & Evans, 2012). She was first presented with two lexigrams, each representing a specific food reward, and allowed to select one. The selected item was then scattered in her outdoor enclosure, along with eight lexigram cards, one of which represented the previously unselected item. Panzee usually first consumed the selected food reward before searching for and returning with the correct lexigram to retrieve the second food reward. Beran et al. (2012) suggest Panzee therefore successfully anticipated her future behaviour, encoding the relevant information needed for the future token retrieval event.

Monkeys have also succeeded during laboratory prospective memory tasks. Evans and Beran (2012) presented capuchins (*C. apella*, n = 9) and rhesus macaques (*M. mulatta*, n = 8) with a computerised task whereby they had to select between a pair of digital stimuli, one of which was rewarded with a food pellet. A prospective memory cue occasionally appeared between trials, signalling to the subjects that a large amount of food pellets would be available at the end of the block of trials, if subjects remembered to select the appropriate icon at the end of the block rather than the icon initiating the next block of trials. Subjects selected the ‘prospective memory’ and ‘next block’ icons when each was the optimal response, with Evans and Beran (2012) concluding they encoded, stored, and responded appropriately to future events. Furthermore, subjects were found to move the computerised cursor to the correct position for the ‘prospective memory’ icon before it appeared, suggesting they anticipated this response, rather than requiring a visual prompt.

During tasks such as these the prospective memory cue is not fully integrated within the task, that is, the stimuli relating to the prospective event is irrelevant to the ongoing discrimination task. Evans et al. (2014) report chimpanzees (*P. troglodytes*, n = 3) succeeded with a prospective memory task embedded within a sorting task. A specific image within the sorting task had previously been prompted as the prospective memory cue, with subjects required to disengage from the task in order to deliver this image to an experimenter in order to obtain a reward. Chimpanzee’s accuracy on the ongoing sorting task varied depending on whether there was a prospective memory requirement or not, with the cognitive control needed to encode and retain the prospective memory resulting in a cost to task performance. Similar results have been found with humans (e.g., Smith, 2003), with Beran et al. (2016a) suggesting the relationship between prospective memory and working memory load may be similar between human and non-human primates.

Defining, and distinguishing between, semantic and episodic future thinking poses a greater challenge. Raby and Clayton (2009) define semantic future thinking as the ability to form scenarios or projections into the future that do not involve personal projection. It is the personal projection into the future event which is said to distinguish semantic from episodic future thinking, whereby individuals themselves are integrated into the imagined future event. Although Suddendorf, Addis and Corballis (2009) propose only episodic prospection confers conscious construction of future scenarios, it is suggested mentally pre-visioning future events draws on both semantic and episodic systems (Irish & Piguet, 2013; Schacter, Benoit & Szpunar, 2017). Martin-Ordas (2016; see also, Osvath & Martin-Ordas, 2014) emphasises the interdependency of the semantic and episodic systems, suggesting future research consider the possibility of these two systems representing two extremes on a continuum. Differences in the complexity of foresights may therefore come from the extent to which episodic and semantic components are incorporated (Martin-Ordas, 2016). Furthermore, semantic processes may provide sufficient flexibility to guide future oriented behaviours (Osvath & Martin-Ordas, 2014).

For example, a series of assessments with western scrub jays (*Aphelocoma coerulescens*) by Clayton and colleagues (see Thom & Clayton, 2016, for a review) have found these birds capable of displaying flexible 'what-where-when' knowledge of future cache retrieval events. Raby, Alexis, Dickinson and Clayton (2007) drew upon aspects of the 'spoon test' proposed by Tulving (2005), reporting evidence of future planning in scrub jays. The spoon test is centred on an Estonian children's story whereby a small girl is disappointed when she cannot eat pudding at a party as she has forgotten to bring a spoon. The memory of this disappointment leads the child to bring a spoon to the next party. Alternating over a period of six mornings jays (*A. coerulescens,* n = 8) experienced two compartments, one where food was always provided, and one where it was not. Following these training trials, the birds were unexpectedly presented with a cacheable food in the evening, and given the opportunity to both eat the food, and to cache it in either of the two compartments. The jays cached significantly more food in the no-food compartment, suggesting they anticipated their future hunger the following morning. A second experiment addressed conditioning accounts of the jays' behaviour. If the jays had learned to associate the separate compartments with different levels of hunger they may have simply cached more food in the compartment associated with hunger, without previsioning future states. Experimental procedure was similar to that of experiment one, however, jays (n = 9) now always received distinct food items in each compartment, for example dog-kibble in compartment A and peanuts in compartment B. When unexpectedly given the opportunity to cache both dog-kibble and peanuts one evening, the jays preferentially cached the food in each compartment that differed to that which they would receive the following morning. Raby et al. (2007) suggest the jays the jays' caching was therefore controlled by anticipation of the following day, rather than an association of each compartment with a particular food type.

Roberts and Feeney (2009a, see also, Suddendorf & Corballis, 2008, 2010) argue these findings do not display evidence of foresight, but may be explainable by semantic processes, with jays acting on factual knowledge rather than a mental representation of the future cache retrieval event. Scrub-jays are scatter-hoarders and may have evolved a tendency to distribute their caches (Roberts & Feeney, 2009b; Suddendorf & Corballis, 2008). The jays may therefore have a predisposition to cache food in novel locations where they have not previously encountered that food. Osvath, Raby and Clayton (2009) counter-argue that caching in novel locations seems less adaptive than provisioning for future needs, and that the use of semantic knowledge does not necessarily preclude the use of future thinking. What is of importance, argue Clayton et al. (2008) are the cognitive processes allowing jays to implement such a heuristic, a process which they propose to involve some form of foresight.

Dufour and Sterck (2008) also drew upon aspects of the spoon test (Tulving, 2005), investigating chimpanzees' (*P. troglodytes*, n = 11) planning abilities during an exchange task paradigm with an experimenter acting as a partner. Subjects were given a ten minute selection period, where a branch, stick, and straw were presented, with the straw the designated objected to be exchanged with the experimenter. Subjects could select any items they wished, and transport them back to a communal room. Following a one hour delay, subjects entered an individual compartment and were offered a food reward in exchange for passing the straw to the human partner. During a pre-test trial, the subjects were not expected to bring straws to the exchange session, as they had not previously experienced the need for straws. The following test trials therefore assessed if chimpanzees would plan ahead, bringing the straws to the exchange session to gain food rewards. The chimpanzees rarely brought the plastic straw into the individual compartment, showing no clear preference for this item over the others presented. These initial results failed to display evidence for future planning.

Despite repeated testing, the chimpanzees failed to select the most reinforced item (the straw), as would be predicted if associative learning alone accounted for their behaviour (Osvath & Persson, 2013). The fact that occasional successes did not lead to future token exchange led Suddendorf and Corballis (2010) to propose behaviour was due to chance, rather than learning or anticipation of future events. Dufour and Sterck (2008) speculate that the requirement to plan what to give, where, when, and to whom, in combination with the cooperation required during exchange tasks may increase complexity. Cooperative contexts have been proposed to be more difficult for chimpanzees than competitive ones (see Hare & Tomasello, 2004), and it is possible that the chimpanzees' initial encounter with the human partner during pre-test trials led them to believe the human was unwilling to cooperate, or a poor quality partner (Dufour & Sterck, 2008). A final experiment therefore presented a tool-using task requiring the chimpanzees to plan which tool to use, when, and where. Tool-using tasks relating to specific material benefits may be less complex than exchange tasks incorporating dynamic social interactions.

Chimpanzees (n = 10) were trained to use a hook shaped tool to retrieve an out of reach bottle containing diluted strawberry juice, in a procedure similar to that employed by Mulcahy and Call (2006, see below). During test trials, subjects were again given a ten minute collection period, where hooks, straws, branches and sticks were all freely available. The subjects entered an individual compartment following a one hour delay, where the juice bottles were presented out of reach. A pre-test trial was again presented, allowing subjects to experience the need for a hook. Of the ten subjects tested in this experiment, three successfully solved the task. Two chimpanzees returned with the hook on 5/17 trials, and one on 7/11 trials, suggesting these individuals anticipated the tool-using task. Dufour and Sterck (2008) consider the findings of this final experiment to demonstrate a sense of the future in chimpanzees. Planning to use a tool in order to obtain a future reward introduces greater cognitive complexity than manipulating a food reward in some way so as to ensure its future retrieval (Mulcahy & Call, 2006). The tool itself has no intrinsic value, other than as a means to retrieve a reward. Despite this increased complexity, during planned tool using tasks the social element is reduced. Requiring subjects to attend to the functional features of potential tools, and consider this functionality for a future point in time may therefore present an appropriate assessment of future thinking in non-human primates.

2.3.3 Planned tool use

Mulcahy and Call (2006) conducted an assessment of future planning using tools with bonobos (*Pan paniscus*, n = 5) and Bornean orangutans (*Pongo pygmaeus*, n = 5), finding subjects capable of selecting and saving tools for future use, both with and without direct visual access to test apparatus, and over delays of up to 14 hours between tool selection and use. Subjects first learned to use a plastic tube to retrieve a food reward, inserting the tube into a cylinder in order to snap a piece of spaghetti that was placed perpendicularly through the cylinder. A bunch of grapes hung on either end of the piece of spaghetti, falling to within subjects reach once the spaghetti had been snapped.

During Experiment 1, subjects entered a test room and were presented with a range of tools. Eight tools were presented, two of which were functional for completing the tube task. The test apparatus remained visible but inaccessible during this five minute selection period. Subjects were then moved to a waiting room for a delay period of one hour, during which the apparatus remained visible. Following this delay, subjects re-entered the test room and were given access to the apparatus, retrieving the food reward if they had selected and transported a functional tool. Subjects successfully retrieved the reward an average of 7/16 trials, with performance per subject ranging from 2-15 successful trials. The subjects took tools to the waiting room on 70% of trials, selecting suitable tools significantly more often than expected by chance. They returned to the testing room with 77.5% of the transported suitable tools.

Experiment 2 followed the same procedure as Experiment 1; however, the delay time was now increased to an overnight period of 14 hours. Of the two subjects tested, the orangutan selected and transported the suitable tool to the waiting room on 11/12 trials. She returned the tool to the testing room and retrieved the reward on seven of these trials. The bonobo successfully returned with the suitable tool on the 8/12 trials that he selected and transported it.

A hook task was used during Experiment 3 , with subjects first learning to use a hook shaped tool to draw in an out of reach bottle of juice suspended from a rope. This experiment assessed whether subjects could select an appropriate tool in the absence of the apparatus, as visual access to the task during tool selection may have cued apes selections. Furthermore, Mulcahy and Call (2006) suggest the use of a liquid reward controlled for the explanation that subjects selected the tool during earlier experiments in response to a current state of hunger, rather than to satisfy a future need. As water was available ad libitum, it is unlikely subjects tool selections were guided by a current thirst. Subjects were again given a five minute selection period, with the hook and three non-functional tools presented. After being relocated to the waiting room, a 1 hour delay was imposed. Subjects then re-entered the testing room and the apparatus was installed, with subjects able to retrieve the reward if they had selected and transported the functional tool. Four subjects were tested, two of which had participated in both previous experiments (*P. paniscus*, n = 2, *P. pygmaeus*, n = 2). Subjects successfully retrieved the reward on an average of 6.5/16 trials. They again transported the suitable tool significantly above chance levels and tended to return the tool to the testing room more often when they had selected functional than non-functional tools.

Mulcahy and Call (2006) note the possibility of the functional tool acting as a conditioned reinforcer during previous experiments, leading to increased selections of this item (see also, Suddendorf & Corballis, 2010). A final experiment therefore aimed to investigate the baseline probability of subjects transporting tools into the testing room in the absence of a tool-using task. The procedure was identical to Experiment 3, subjects were rewarded for bringing the correct tool to the testing room; however, no apparatus was presented. Two orangutans and two bonobos were tested, none of which had participated in previous experiments. Subjects performed significantly poorer than those in

Experiment 1 or Experiment 3, with only two subjects returning with the correct tool. Overall subjects successfully selected and transported the suitable tool on an average of 1.8 trials. Performance in Experiment 4 suggests subjects' selections were not based on simple associations, as the tool was still reinforced and subjects performed poorly across multiple trials despite earlier success. Mulcahy and Call (2006) conclude that subjects displayed future planning, selecting and transporting tools based on future needs rather than current ones, however alternative explanations have been proposed. Of the four subjects participating in Experiment 4, two never returned with the appropriate tools. These subjects therefore had no opportunity to learn the contingency between tool return and reward acquisition. The remaining two subjects returned tools at a rate similar to those in previous experiments. Thom and Clayton (2016, see also, Suddendorf & Corballis, 2010) suggest performance in earlier tasks may therefore be attributable to conditioning, with subjects forming associations between the tool and reward. Furthermore, despite the use of a liquid reward during Experiment 3, the lack of control over subjects' motivational states has led to criticism. Suddendorf and Corballis (2007) suggest these findings are explainable by the Bischof-Köhler hypothesis as subjects' behaviour may have been driven by current desires rather than foreseeing of future states.

Osvath and Osvath (2008) sought to further the findings of Mulcahy and Call (2006), conducting a series of experiments to investigate foresight in chimpanzees (*P. trogylodytes*, n = 2) and a Sumatran orangutan (*P. abelii*, n = 1). The series of experiments included a baseline assessment of selection of functional tools for future use, a self-control test, an association control, and finally a test of pre-experience of the particularities of a future reward retrieval event. The apparatus used throughout consisted of a transparent bottle housed within a wooden box with a clear panel on one side, allowing visual access to the contents of the bottle. The bottle was filled with fruit soup, extractable by inserting a straw into a hole on the top of the box. Subjects received a tool-use training phase, where they were first shown the function of the straw before being handed it and allowed to retrieve the reward. Following this, subjects had access to a reward room containing the apparatus, with no tools available, allowing them to experience the need for a tool. The following day, subjects were presented with a tray containing the functional straw tool and three non-functional objects whilst the apparatus was in view. Subjects were permitted to make one selection, being allowed to enter the reward room after a delay of one hour. All subjects selected the functional tool during this training trial.

Test trials during Experiment 1 followed a similar procedure, assessing subjects' baseline performance. Subjects entered a selection room, were permitted to make one selection whilst the apparatus was out of view, and then returned to their main enclosure with their selected item. A delay period of 70 minutes took place before the apparatus was installed and subjects gained access to the reward room, retrieving the reward if they had selected and transported the functional tool. Each subject received 14 trials, with two subjects selecting the functional tool across all trials, subsequently returning with the tool and retrieving the reward on 11 trials. The remaining subject selected the functional tool on 13 trials, successfully retrieving the reward on 12 of these trials.

These results are consistent with those of Mulcahy and Call (2006) in displaying apes ability to select and save tools for future use; however, Osvath and Osvath (2008) suggest they further these findings in several respects. Whilst Mulcahy and Call (2006) blocked subjects access to the reward during tool selection, Osvath and Osvath's (2008) procedure required subjects to make selections in a separate location, with no cue to the presence of the apparatus. Furthermore, the subjects could witness the 'absence' of the apparatus prior to entering the selection room, leading Osvath and Osvath (2008) to suggest the results are indicative of a greater cognitive detachment from the perceptual stimuli. The apes also returned to their social groups during the delay period, rather than being detained in a waiting room with their selected tool. This likely increased the cognitive load during the delay, distracting attention from the selected tool and reducing the likelihood of the tool cueing a desire for the reward. In a critique of these experiments, Suddendorf, Corballis and Collier-Baker (2009) suggest the apes initial training trial may have resulted in a strong positive association between the straw and reward. Subsequent selections of the functional straw may have been guided by this associative learning, in response to an immediate desire for fruit soup, rather than by anticipating future hunger. Osvath (2010) responded to this critique, noting that one-trial learning was an integral feature of the methodology and that it does not equate to associative learning. If associative learning had occurred, whereby the apes selected the straw in anticipation of immediate reward retrieval, this behaviour would be expected to become extinguished over future trials encompassing a delay between tool selection and reward retrieval.

Experiment 2 introduced an assessment of self-control, investigating subjects’ ability to inhibit a current desire, and decline a small reward, in order to obtain a tool usable to retrieve a greater future reward. Self-control tasks set present and future needs in opposition by design, representing an alternative assessment of the Bischof-Köhler hypothesis (Osvath & Osvath, 2008; Thom & Clayton, 2016). The experimental procedure was similar to that in Experiment 1; however, a favourite fruit was now added during the selection process, with one selection again permitted per trial. The orangutan selected the functional tool on 9/14 trials, returning with it in during all trials. One chimpanzee selected the straw on eight trials, returning with the tool on four of these trials. The final chimpanzee chose the functional tool during 11 trials and returned with the tool on nine trials. The apes thus showed the ability to override current needs in favour of satisfying a different future need. Osvath and Osvath (2008) note the selection of the tool does not lessen a current desire for fruit soup, nor does it remove a current desire for the offered favourite fruit, and would therefore be a maladaptive choice without reference to future needs.

If the subjects ascribed the tool an intrinsically high value, in isolation from its future function, this may have guided selections across Experiments 1 and 2, rather than planned future use. Osvath and Osvath’s (2008) third experiment therefore aimed to control for associative learning explanations. The experimental procedure was similar to that in previous experiments, with subjects first choosing one item from a selection of one functional and three non-functional tools. Following this, they were immediately presented with the same choice (including a duplicate of their previously selected tool), with the addition of a piece of preferred fruit. All subjects chose the fruit during the second selection across all 14 trials, suggesting they saw the tool as a means to an end, selecting it for future use (if they recognised an opportunity to use it to obtain a reward later), rather than due to positive associations. Osvath and Osvath (2008) further note this does not exclude any associations between the tool and the reward, merely that this association was not an arbitrary one expected from simple associative learning accounts. Suddendorf, Corballis and Collier-Baker (2009) argue that the apes' selection of the straw during the first presentation may have been guided by an expectation of immediately receiving the reward, and after having obtained the functional tool, they then chose the food reward. This explanation seems unlikely, given that the apes had never experienced reward presentation directly following tool selection (Osvath, 2010; Thom & Clayton, 2016). Furthermore, Osvath (2010) notes that Suddendorf, Corballis and Collier-Baker's (2009) critique acknowledges the apes understanding of the straw as a functional tool and of its anticipated use, rather than assigning it an arbitrary intrinsic value, which is what this experiment aimed to assess.

In the final experiment, subjects selected one item from a range of three novel items and one familiar item, with one of the novel items usable as a tool to obtain a future reward. The familiar item had previously been associated with food retrieval, but was non-functional with the current apparatus. The range of functional tools presented differed from the original functional tool in various features (such as length, flexibility, and shape), with the non-functional items often sharing characteristics with the previously presented functional straw. This experiment therefore aimed to assess if subjects were pre-experiencing the future reward retrieval, attending to the particularities of the presented objects in order to select the item that would be functional at a later time. Over the 12 trials, the orangutan selected the functional tool in 11 trials, one chimpanzee in ten trials, and the final chimpanzee in nine trials. As a group the apes selected the functional tool significantly above chance levels. They confidently used these novel functional items to retrieve the reward after returning to the reward room suggesting they had attended to the functionally relevant features of the items during selection. Osvath and Osvath (2008) consider the applicability of semantic prospection in accounting for these findings. As subjects were capable of selecting novel items based on particular features of functionality, their behaviour may be more readily explained by mentally pre-experiencing future situations, comparing items functionality to the future reward retrieval event in each instance. Despite offering associative-learning as an explanation for previous experiments, Suddendorf, Corballis and Collier-Baker (2009) concede that the results of this final experiment suggest the apes learned relevant affordances rather than simple perceptual stimulus-response relations. Furthermore, they note this is unsurprising, given apes relatively poor ability at learning arbitrary associations. It seems implausible though to attribute the ape's performance during Experiment 4 to advanced representational skills (due to poor associative-learning abilities), but to offer associative learning as an explanation for previous experiments (Osvath, 2010).

Overall, these findings provide support for the ability to attend to functionally relevant features of tools and plan for future use in large apes. Findings are not easily explained by associative learning accounts, and Osvath and Osvath (2008) suggest subjects selections were unlikely guided by current motivations, as predicted by the Bischof-Köhler hypothesis. If subjects were incapable of previsioning future desires, they would be expected to select the immediately available fruit during the second experiment to satisfy immediate desires. The possibility that the presence of the functional tool created an immediate desire for fruit soup, and subjects therefore responded to competition between two present desires is potentially negated by the findings from Experiment 3. However, Suddendorf, Corballis and Collier-Baker (2009) remain unconvinced that this assessment contests the Bischof-Köhler hypothesis as it does not provide evidence for future need anticipation; the apes may have learned there is a future need for the straw, without any awareness of differing future needs. Furthermore, Roberts and Feeney (2009a) suggest the apes may have based selections on functional properties, again without anticipating the straws future use. Bräuer and Call (2014) sought to address these criticisms by presenting apes with a tool manufacturing task. If apes simply selected tools based on their functional properties, they would not be expected to construct tools, as the materials presented to them were non-functional. Furthermore, if the apes were incapable of referencing future needs, they may produce the same number of tools regardless of how many the future task required.

Twenty five apes (*P. troglodytes*, n = 14, *P. paniscus*, n = 4, *Pongo* sp., n = 7) were presented with a task containing eight Plexiglas tubes, mounted next to one another, perpendicular to the enclosure mesh. The tubes were presented outside the enclosure; however, apes had access to the open end of the tubes. When baited with a food reward, subjects could insert a stick into the tube, pushing the reward along to a hole at the opposite end. The reward then fell to within subjects reach. The apes were presented with soft wooden boards which they could split into strips. These boards were short enough that, once pushed into the tube, a stick could not be retrieved. Subjects were therefore required to construct multiple stick tools in order to retrieve all the rewards. Training trials allowed subjects to familiarise themselves with the tool making procedure and apparatus, with one tube being baited and one wooden board provided. Subjects who passed two training trials passed to test trials, which consisted of two phases. An eight minute tool preparation phase presented the subjects with two wooden boards with visual access to the apparatus. Following this, the apparatus was unblocked and a tool use phase occurred for either 50 or 120 seconds. The wooden boards were not removed prior to the tool use phase, meaning subjects could continue to make 'late tools' during this period. Of the 25 subjects, 12 received the 50 second tool use phase only, three subjects the 120 second phase only, and the remaining ten subjects both time delays. There were three different conditions within the tool use phase; either all tubes baited, one tube baited, or empty apparatus whereby eight grapes were placed in the catchment area under the tubes within subjects' reach. The empty apparatus condition therefore did not require subjects to use any tools. Each subject received 15 trials on each condition.

There was a significant difference between conditions in both the 50 second and 120 second tool use phases, with subjects preparing tools in more of the 'all baited' trials than the 'one tube' or 'empty' apparatus conditions. Despite manufacturing less than two tools per trial on average, subjects also produced a greater number of tools in the 'all baited' condition across both time periods. The production of tools prior to the tool use phase did not increase subjects' success at reward retrieval during the 50 second phase. Bräuer and Call (2014) suggest this supports tool manufacture being guided by the future task, rather than reinforcement for tool production. The findings do not represent planning for a drive state not currently experienced, as the apparatus was present and baited during tool manufacture. However, the apes' differential tool production did relate to differing future needs, in contrast to Roberts and Feeney's (2009a) proposal that apes select tools on the basis of their functional properties, rather than future use.

A comparison between phases for those subjects who participated in both the 50 second and 120 second trials found little difference in tool preparation, suggesting greater time pressure did not lead to increased tool manufacture. Bräuer and Call (2014) propose two possible explanations for this. The apes may not have needed to manufacture more tools in advance as they continued to make tools during the tool using phase. Alternatively, anticipating the quantity of required tools may represent a limit in apes planning ability, a proposal that requires further investigation. Investigation of planned tool use in monkeys has provided more equivocal results, with Dekleva, van den Berg, Spruijt and Sterck (2012) reporting long tailed macaques (*Macaca fascicularis*) failed to transport tools for future use without extensive shaping of their behaviour. Experiment 1 in the first set of experiments aimed to replicate previous findings described above (Mulcahy & Call, 2006; Osvath & Osvath, 2008), presenting macaques (*M. fascicularis*, n = 6) trained in using a simple raking-in tool with a choice between two tools and two pairs of distractor objects presented on a table outside the selection compartment. Subjects could pull the selected item into the enclosure through the mesh. In order to succeed, macaques were required to select the functional tool, transport this tool first into a waiting compartment for a delay period of five minutes, and then to the test compartment where they could use it to obtain a food reward. Dekleva et al. (2012) note the reward table was not visible from the tool selection site and subjects were not hungry at time of tool selection. Therefore, it is proposed that selection and transportation of the functional tool would display anticipation of future needs.

Monkeys never transported tools to the test compartment. Despite occasional selections of the tool, they predominantly selected distractor objects. These distractor objects were sometimes transported, but never used to attempt to retrieve rewards. As an alternative explanation to the inability to anticipate future needs, Dekleva et al. (2012) suggest the monkeys may have failed to connect the temporally and spatially distant events of object selection and reward retrieval. A further two experiments therefore manipulated the experimental set-up, making the task easier by gradually reducing the space and time between these two events. Monkeys continued to fail to select and transport the tool when shown the baited table immediately prior to selection, a manipulation intended to cue the monkeys' memory of previous failures. They also failed to transport tools when the delay period was removed and functional tools were provided inside the waiting compartment. In this experiment monkeys were also free to return to the waiting compartment to retrieve the tools after experiencing the need for them in the testing compartment. A final test found the monkeys had a context specific preference for the tool. The test measured subjects' responses to the previously presented tools and distractor objects across three different contexts. The objects were presented on the table outside the selection compartment, and on both a baited and non-baited table outside the test compartment. The monkeys displayed a shorter latency to initial contact, and longer manipulation time when the tools were presented in the context that they had originally been trained to use them in (i.e., rake tools on the baited table). Dekleva et al. (2012) consider the effect of training on subjects' behaviour, noting that the monkeys required many sessions to learn to reposition the tool in order to rake in the reward. This training may have inhibited subjects learning different manipulations of the tool outwith the original context.

Subjects (n = 5) were then trained to transport the rakes in incremental stages, with rewards given at each stage if subjects displayed the appropriate behaviour. Stage one required manipulation of a tool partially inserted into the enclosure onto the test table to rake-in a reward. The next stage required transporting the tool with the overall distance to be travelled being gradually increased until subjects transported tools from the waiting compartment to the test table. Finally, the monkeys were required to transport tools from the selection to the test compartment, with the distances again being gradually increased. Breaking up the sequence of required actions into smaller rewarded movements thus allowed the monkeys to learn the required behaviour. A second series of experiments presented a similar procedure to that in the first experiment; however, two tools were now inserted into the selection compartment. Incrementally increasing delay periods (5 minutes, 10 minutes, 20 minutes) were imposed before access to the test compartment was given. The five subjects transported the rake to the test compartment in 53% of completed five minute delay trials. Two subjects were tested at the ten minute delay period, one transported the rake on 12/12 trials, with the other individual succeeding on 4/5 trials. At the 20 minute delay period, one subject failed to transport the tool in any trials. The other succeeded in 3/4 completed trials, completing the trained sequence of behaviours despite the 20 minute interval. Although monkeys learned some skills with incremental training, their performance was poorer than large apes at transporting functional tools for future use. Osvath and Persson (2013) speculate that there may differences in the cognitive or learning systems underlying future planning in monkeys and large apes. Monkeys failed to plan even on simplified versions of a task that apes succeed in, only displaying the appropriate behaviour after learning the required sequence of actions by progressive step-wise training.

2.4 Overview of the current thesis

This thesis aims first to assess whether gibbons are capable of reliably attending to the relevant features of two rakes during a zero-order raking in task, discriminating between a functional and a non-functional rake. Gibbons have previously succeeded during zero-order raking in tasks, suggesting an understanding of the relationship between the manipulable object and a food reward (Cunningham, Anderson & Mootnick, 2006, 2011; Inoue & Inoue, 2002). However, these assessments do not provide evidence of whether gibbons understand which properties of the rakes make them functional for reward retrieval, or whether they can discriminate between objects which are functional for reward retrieval and those which are not. There are few reports of tool use in wild gibbon populations (see Shumaker et al., 2011); however, typically non-tool using species of monkey have displayed proficiency at this task (Cummins-Sebree & Fragaszy, 2005; Hauser, 1997; Santos et al., 2006; Spaulding & Hauser, 2005). Spaulding and Hauser (2005) proposed non-tool using species have an innate capacity for recognising the functionally relevant features of objects, given sufficient task relevant experience. If this hypothesis is correct, it was expected gibbons would also display the ability to attend to specific perceptual features related to functionality.

Pilot testing (Chapter 3) assessed gibbons’ innate understanding of what makes an object functional in reward retrieval during a raking in task similar to the inverted rake task developed by Povinelli and Reaux (2000). Two rakes were presented, one functional and one non-functional, with a food reward positioned in front of both. This assesses whether gibbons can reliably select the functional rake when required to take into account the three dimensional relationship between rakes and a food reward. Tasks of this sort potentially present an increased cognitive difficulty than tasks where the reward is positioned either in-front of a tool or not (as in previous research with New World monkeys, see Cummins-Sebree & Fragaszy, 2005; Hauser, 1997; Santos et al., 2006; Spaulding & Hauser, 2005). Gibbons performed poorly on this task, potentially as a result of difficulty discriminating between the rakes, given their similarity in appearance. Chapter 4 therefore assesses gibbons' ability to select a functional rake when presented with two rakes with more obvious perceptual differences. A series of experiments are presented, assessing gibbons' ability to generalise learning of functionality across different rake sets. Furthermore, this chapter presents experiments both requiring consideration of whether the reward is positioned in-front of the rake or not, and a repeat of the task used in pilot testing, assessing if the experience gained through experiments with more marked perceptual differences in the rakes leads to improved performance with this apparatus.

Martin-Ordas (2016, see also, Osvath and Martin-Ordas, 2014) argues that focus on a dichotomous framework of prospective cognition (i.e., episodic or not) has limited assessment of the range of prospective abilities found in non-human primates. In line with this argument, Chapters 5 and 6 present assessments of basic prospection and component processes within gibbons. Chapter 5 investigates whether gibbons can select un-baited functional rakes in order to retrieve a food reward delayed up to a period of five minutes. This task is similar to that presented by Mulcahy and Call (2006) and Osvath and Osvath (2008) assessing planned tool use in large apes. However, given gibbons failure at tasks incorporating true tool use (Cunningham, 2006), subjects were not required to transport or properly orient the rakes during this task. The task presented in Chapter 5 therefore assesses whether gibbons can connect the temporally and spatially distinct events of rake selection and reward retrieval, as has been found within monkeys following progressive training (Dekleva et al., 2012). Chapter 6 presents a preliminary investigation into self-control capacities within gibbons. The ability to display self-control is a crucial component of prospective cognition; regardless of how advanced other cognitive mechanisms are, actions towards future outcomes cannot be taken unless an individual can first inhibit current desires in favour of future ones. The series of experiments in Chapter 6 assess whether gibbons are capable of foregoing an immediately attainable small reward in order to select a functional rake, useable for retrieval of a large reward. Finally, Chapter 7 presents a general discussion of the main findings, including genera, sex, and age group differences across the presented experiments, before considering how the current research furthers understanding of both basic prospection and self-control abilities within the small apes.

Chapter 3: Pilot testing – Selection of functional rakes

3.1 Introduction

Gibbons (*Hylobatidae*) are not prolific tool users; the few reports of tool use in wild gibbons consist of breaking off and then dropping branches, and shaking branches during vocalisations (see Shumaker, Walkup & Beck, 2011, for a detailed overview). A greater diversity of tool using capacities has been reported in captive gibbons (see Chapter 1, section 1.4, for more details). Early investigations found captive gibbons' exploratory behaviour with unfamiliar objects to be less extensive than that displayed by large apes but more diverse than that reported for monkeys (Bernstein, Schusterman & Sharpe, 1963; Parker, 1974). Parker (1974) proposed the larger relative brain size of the apes may allow for a greater behavioural diversity, which in turn may lead to improved problem solving skills as a result of the increased repertoire of available behaviours.

Gibbons have displayed proficiency at problem solving tasks presenting directly perceivable relationships between manipulable objects and food rewards (see Chapter 1, section 1.4.2, for more details). Beck (1967; see also, Köhler, 1925) found gibbons performed comparably to chimpanzees (*Pan troglodytes*) during a series of string pulling problem solving tasks, varying from simply pulling in a food reward attached to a piece of string, to more complex manipulations. Gibbons have also succeeded at zero-order manipulation tasks requiring them to use a rake shaped object to draw in a food reward (without the need to reorient the rake), and perform at a level comparable to more frequent tool using species (chimpanzees) during second-order trap-table tasks, suggesting an understanding of the causal relationships between three factors; the rakes, food reward, and a trapping hole (Cunningham, 2006; Cunningham, Anderson & Mootnick, 2006, 2011; Inoue & Inoue, 2002). However, it remains unclear whether gibbons are capable of demonstrating an understanding of which features make an object functional for reward retrieval during these types of tasks.

Both tool using and typically non-tool using species of New World monkeys have demonstrated a level of understanding of the properties of objects that make them useful for reward retrieval during zero-order manipulations (see Chapter 2 for more details). Cotton-top tamarins (*Saguinus oedipus*), for which there are few reports of spontaneous tool use, and capuchins (*Cebus apella*), who are habitual tool users (see Shumaker et al., 2011, for a review), were found to attend to functionally relevant features of raking in tools, selecting rakes with the best functional design when presented with a choice between two rakes differing in perceptual features such as size, texture, shape and colour (Cummins-Sebree & Fragaszy, 2005; Fujita, Kuroshima & Asai, 2003; Hauser, 1997; Hauser, Pearson & Seelig, 2002). Subjects selected rakes with irrelevant featural changes such as colour, over those with changes that impacted on functionality (e.g., shape). These findings led Hauser et al. (2002) to propose the ability to differentiate between functionally relevant features of potential tools was an innate capability, not limited to tool-using species.

Spaulding and Hauser (2005) revised this hypothesis, finding experience was needed in order for tamarins (*S. oedipus*) and common marmosets (*Callithrix jacchus*) to reliably attend to functionally relevant object features during a raking in task. Performance was initially poor during a partial replication of Hauser and colleagues' earlier research (Hauser, 1997; Hauser et al., 2002), which presented subjects with two rakes varying in perceptual features such as colour, shape and size. Neither species reliably selected the rake with perceptual changes that did not affect functionality. However, during further testing incorporating a range of novel tools, subjects' performance reached significance. Spaulding and Hauser (2005) concluded these species may be endowed with an ability to recognise functionally relevant perceptual features, but task relevant experience is needed in order to achieve reliable performance. This hypothesis is supported by Santos, Miller & Hauser's (2003) research, using an expectancy violation procedure to assess whether previous object manipulation experience is required in order for subjects to attend to the functionally relevant features of tools which they are not directly manipulating. Subjects looking time was measured as they witnessed an experimenter using a tool to manipulate a reward, with the perceptual features of this tool varied across experiments. Tamarins (*S. oedipus*) with prior object manipulation experience attended to both functionally relevant shape and orientation changes whereas experimentally naïve rhesus macaques (*Macaca mulatta*) attended to shape changes, but not functionally relevant orientation changes.

The current research aimed to assess if gibbons could attend to the functionality of two different rakes during a zero-order raking in task (see section 3.2.2). Given the support for this ability in New World monkey species, and gibbons' phylogenetic position located between the large apes and monkeys, it was predicted they would be capable of attending to the relevant features of rakes that make the object functional in reward retrieval given sufficient experimental experience. As gibbons have shown more diverse manipulative behaviour with novel objects than monkeys, and performed comparably to chimpanzees (*P. troglodytes*) in research involving object manipulation to attain a reward (Beck, 1967; Cunningham, 2006; Cunningham et al., 2006), it was predicted that these apes would understand what properties of objects make them useful in goal attainment.

The task used in the current research was similar to that developed by Povinelli and Reaux (2000) to assess spontaneous understanding of causal relationships in chimpanzees (*P. troglodytes*): this inverted rake task involves consideration of the three-dimensional relationships between the tools and reward, taking in to account the rake orientation rather than simply whether the reward is positioned in front of the rake. Povinelli and Reaux (2000) presented chimpanzees with two rakes, one oriented so it was functional for reward retrieval, and one in a non-functional orientation. The rake ends consisted of two vertical prongs attached to either end of a horizontal bar and the rake handle attached perpendicularly to the middle of this bar. Thus, the rakes were functional when laid with the horizontal bar on the testing platform, and non-functional when inverted, so the rake stood on the two vertical prongs with the reward passing through the gap underneath.

Povinelli and Reaux (2000) found chimpanzees selected the functional rake on 57% of trials during an initial presentation of this task consisting of four trials per subject. Conducting a replication of Povinelli and Reaux (2000) with two species of monkey, Santos, Pearson, Spaepen, Tsao and Hauser (2006) found tamarins (*S. oedipus*) selected the correctly oriented rake on 52% of the six presented trials, with vervet monkeys (*Cercopithecus aethiops*) performing slightly better, selecting the correct rake on 60% of trials. The findings suggest subjects displayed some understanding of the functionally relevant features; however, performance did not differ greatly from chance levels. Povinelli and Reaux (2000) and Santos et al. (2006) presented a limited number of trials in order to assess subjects spontaneous understanding of the task, and it is possible with further testing these species would have learned to attend to the relevant rake features. In the current research, it was decided to increase the amount of time available for learning by presenting a greater number of trials to subjects, assessing whether this increased experience would lead to task sufficient performance. Given the relatively poor performance of chimpanzees and monkeys on this task, it was not expected gibbons would display a spontaneous understanding of the functionally relevant features of the rakes. However, in line with Spaulding and Hauser's (2005) proposal that non-tool using species may recognise functionally relevant features of objects given sufficient task relevant experience, it was expected gibbons would learn to differentiate between the rakes over an extended testing period.

The current research presented a greater difference in appearance between the two rakes than that used during Povinelli and Reaux's (2000) inverted rake task, in an effort to make the differences between the rakes more obvious (see section 3.2.2) given the reported poor performance on cognitive tests in gibbons as a group. The non-functional rake was shaped similarly to the 'inverted non-functional rake' used by Povinelli and Reaux (2000) with chimpanzees (see Figure 3.4). However, the functional rake end was a solid rectangular shape, rather than the same 'two-pronged functional rake' presented with the horizontal bar on the platform and prongs projecting upwards (see Figure 3.3). The task therefore presented one functional and one non-functional rake, baited with food rewards and placed on a platform outside subjects' enclosures (see Figure 3.5). Subjects were required to select one rake and pull it towards them, sliding the food reward to within reach if the functional rake was selected (the food reward would not move closer if the non-functional rake was chosen although the rake could still be pulled in towards the subject). As noted previously, raking in tasks of this nature allow for investigation of what subjects understand about the relationship between the rake and goal object

(Povinelli & Reaux, 2000). In line with Shumaker et al.'s (2011) definition, this task does not represent true tool use, as subjects were not required to orient the rakes in the correct position before pulling in. However, as the focus of the study was to investigate subjects' ability to attend to functionality, and gibbons have shown previous success at raking-in objects (Cunningham, 2006; Cunningham et al., 2006, 2011; Inoue & Inoue, 2002), this was considered to represent an appropriate task.

3.2 Methods

3.2.1 Subjects

Testing was conducted at the Gibbon Conservation Center (GCC), a non-profit organisation located in California (USA), which aims to promote the conservation, study and care of gibbons (Gibbon Conservation Center, 2018). Approximately 40 gibbons were housed at GCC, including representatives from five species and all four genera. Gibbons were fed eight times throughout the day with a variety of fruits, vegetables and primate biscuits, with water available ad libitum. The gibbons were housed in outdoor enclosures, consisting of both a primary (average 10 x 3 x 4 metres high) and smaller secondary (average 4 x 3 x 2.5 metres high) section that was routinely available at all times but could be partitioned off to separate individuals. Where possible, gibbons were housed in groups that approximated natural family groups or breeding pairs (see Table 3.1 for details), and received minimal human interaction other than for feeding and medical needs. In order to simulate a housing situation similar to that found in natural settings where gibbon groups often have auditory but limited visual contact with each other, tarpaulins and vegetation were used to form visual barriers between enclosures, minimising direct visual contact between groups (see Figure 3.1). Enclosures were furnished with branches and slacklines, floor level feeding platforms, raised feeding trays or buckets, and raised sleeping areas with insulated shelters (see Figure 3.2).

Figure 3.1: Example of an enclosure at the Gibbon Conservation Center (GCC) with vegetation and tarpaulins providing visual barriers between groups.

Figure 3.2: Example of an enclosure at GCC furnished with slacklines and branches and an insulated shelter on the raised sleeping platform at the rear of the enclosure.

Subjects were 31 gibbons housed at GCC, including nine eastern hoolock (*Hoolock leuconedys*), nine javan (*Hylobates moloch*), six northern white cheeked (*Nomascus leucogenys*), five pileated (*Hylobates pileatus*), and two siamang gibbons (*Symphalangus syndactylus*) (see Table 3.1). A number of gibbons had previously participated in cognitive testing. In particular, some individuals had completed tests using a similar raking in task between 2002 and 2009 (see Cunningham, 2006; Cunningham et al., 2006; 2011). However, to our knowledge, subjects had no other exposure to object-use tests, and in general cognitive testing of gibbons at GCC has been minimal. Testing took place during week days when GCC was closed to the general public, between the hours of 7am and 12 noon when gibbons were most active. Subjects were tested in their home enclosures with water freely available and no disruption to their normal feeding schedule. Of the 31 subjects included in testing, 25 completed trials with the remaining 6 showing no interest in the task. These subjects failed to select a rake on any presentation and were therefore excluded from further testing. Vok (enclosure 13a) and Hmawe Ni (enclosure 7) became disinterested in the task and were excluded from analyses as they completed only 19 and 20 trials, respectively.

Subjects were tested individually, with the exception of those in enclosures 2, 3, 5, 11, and 13a. Individuals in these enclosures, housing subjects from all species except *S. syndactylus*, became anxious when separated. Although separation would have been preferable, this was balanced with welfare issues caused by separation and so the decision was made to test these individuals in their groups. Neither of the individuals in enclosure 3 showed interested in the task, failing to complete any trials. In enclosure 2, only one individual (Khusus) displayed interest in the task. Khusus was therefore tested in her family group, consisting of one adult male and two offspring, none of whom participated in testing. Three subjects (Domino, Iszie and Tuk) in enclosure 5 completed testing; the remaining infant in the group was too young to be included in testing. During trials the apparatus remained out of reach until the current focal individual approached the apparatus. Subjects in this enclosure therefore did not receive additional trials that could have impacted their learning whilst testing was completed with all individuals, although we cannot discount the possibility that seeing another group member tested could have impacted performance (see Whiten & van de Waal, 2018, for a review of social learning in primates). Phy Gyi and Arthur were separated for testing in Enclosure 11; however, this pair was housed with two infants who were too young to participate, with the youngest staying with Phy Gyi during testing and the older infant remaining with Arthur.

Enclosures 9 and 13 were partitioned into two separate areas for the duration of the testing period. Truman and Violet were therefore tested individually (enclosure 9). One subject in enclosure 13 (Paddy) was housed alone in the secondary area (enclosure 13b) and was therefore tested individually. There were five individuals housed in enclosure 13a, with two infants too young to be included in testing. The remaining three individuals were initially tested in the same manner as enclosure 5, with the apparatus only being placed within reach when a focal individual approached to avoid subjects receiving additional trials. However, Canter did not complete any trials and Vok became disinterested in the task and was excluded from further testing, with only Ricky reliably taking part in testing. With the exception of U Maung Maung (enclosure 7) and Khusus (enclosure 2), who were tested in their family groups but were the only individuals in these enclosures who participated, and Reg (enclosure 4), who was housed individually, subjects may therefore have witnessed other gibbons in their enclosures completing testing, potentially giving them a learning advantage. However, being the second, or third, individual tested in an enclosure did not appear to improve performance levels (see 3.3.1.1 for analysis of order effects). The order in which the individuals in each enclosure were tested (where more than one individual in the enclosure participated) was randomised, so the same individual was not tested first each day.

Table 3.1: Subjects' species, age, sex, housing information, whether subjects had previous experience with raking-in tasks, and whether they completed testing, or the reason why they were excluded.

Subject	Enclosure	Species	Age	Sex	Previous Experience[1]	Group	Completed testing
Chan Thar	6	*H. leuconedys*	9	F		Male/female pair	✓
Win Bo	6	*H. leuconedys*	12	M		Male/female pair	✓
Hmawe Ni	7	*H. leuconedys*	11	F		Male/female pair	Became disinterested in task.
U Maung Maung	7	*H. leuconedys*	14	M	✓	Male/female pair	✓
Betty	10	*H. leuconedys*	16	F	✓	Female & male/male siblings	✓
Khin Maung Win	10	*H. leuconedys*	7	M		Female & male/male siblings	✓
U Myint Swe	10	*H. leuconedys*	5	M		Female & male/male siblings	✓
Arthur*	11	*H. leuconedys*	19	M	✓	Family	✓
Phy Gyi*	11	*H. leuconedys*	12	F		Family	✓
Chloe	1	*H. moloch*	24	F	✓	Family	✓
Ivan	1	*H. moloch*	41	M	✓	Family	✓

Khusus*	2	*H. moloch*	20	F	✓	Family	✓
Oula*	2	*H. moloch*	6	F		Family	Did not approach apparatus.
Shelby*	2	*H. moloch*	31	M		Family	Did not approach apparatus.
Perak*	3	*H. moloch*	13	M		Family	Did not approach apparatus.
Simpang*	3	*H. moloch*	14	F		Family	Did not approach apparatus.
Reg	4	*H. moloch*	14	M		Solitary	✓
Medina	8	*H. moloch*	12	M		Solitary	Did not approach apparatus.
Canter*	13a	*N. leucogenys*	6	M		Family	Did not approach apparatus.
Ricky*	13a	*N. leucogenys*	29	F	✓	Family	✓
Vok*	13a	*N. leucogenys*	31	M	✓	Family	Became disinterested in task.
Paddy	13b	*N. leucogenys*	10	M		Solitary	✓
Asteriks	15	*N. leucogenys*	15	F		Family	✓
Lucia	15	*N. leucogenys*	5	F		Family	✓
Domino*	5	*H. pileatus*	19	M		Family	✓
Iszie*	5	*H. pileatus*	3	F		Family	✓
Tuk*	5	*H. pileatus*	21	F	✓	Family	✓
Truman	9	*H. pileatus*	11	M		Solitary	✓
Violet	9	*H. pileatus*	5	F		Solitary	✓
Karenina	14	*S. syndactylus*	15	F		Mother/daughter	✓
Marlow	14	*S. syndactylus*	9	F		Mother/daughter	✓

✓ = Completed testing

* Not separated for testing

[1] These subjects participated in previous research incorporating a similar zero-order raking in task (see Cunningham, 2006; Cunningham et al., 2006, 2011).

3.2.2 Experimental procedure and apparatus

The main task used throughout testing required subjects to pull in a rake to slide a food reward within reach. To allow subjects to learn about rake functionality, they were presented with one functional and one non-functional rake, both baited with a preferred food reward (blueberries). This food item was selected as it was not normally included in the gibbon's diet, but was reported as a preferred food by the GCC staff. The rake handles were constructed from black polypropylene rods, 60cm in length, with a 6cm diameter. The length of the rake handles meant gibbons could not reach the food reward without first pulling in the rake. A small bolt inserted through the end of the rake handle elevated the handle from the platform slightly for easier grasping by the gibbons elongated hands. Rake ends were made from a wood/plastic composite (60% timber, 40% polyethylene), with the functional rake (F1, see Figure 3.3) having a 34cm by 10.5cm rectangle end to act as a rake. The non-functional rake (NF1, see Figure 3.4) had a rectangle-shaped end, 34cm by 10.5cm, with a smaller rectangle cut out the middle (30cm by 6cm), leaving two legs of 2cm on either side and rendering it useless for reward retrieval. If pulled, the non-functional rake simply moved over the reward as the food went through the cut-out section; however, the overall shape was consistent with the functional rake.

Figure 3.3: F1 functional rake end.

Figure 3.4: NF1 non-functional rake end.

The rakes and food reward were presented on a platform (180cm by 75cm) situated outside subjects' enclosures, long enough that rakes could be positioned with the handle ends initially out of direct reach of the subjects. For enclosures where individuals were separated for testing, the position of this platform varied dependent on whether the separated focal individual was located in the primary section of the enclosure, or the smaller secondary area. For enclosures were only one individual participated, the position of the platform was kept consistent throughout testing. Two strips of wood (100cm by 2cm) ran down either side of the platform, with a third acting as a dividing partition along the centre. These prevented the reward from either rolling off or passing from one side of the platform to the other (see Figure 3.5). The platform was presented to all subjects at a raised level (90cm), as some subjects were not comfortable coming down to floor level. This height was consistent with the level of feeding buckets used by most gibbons. All apparatus were cleaned with non-toxic cleaning agents between presentations at different enclosures.

Figure 3.5: Functional and non-functional rakes positioned on testing platform. The non-functional rake remains in the starting position, with a Javan gibbon (*H. moloch*) selecting the functional rake.

A trial began with the experimenter baiting both rakes in full view of the subject. A five second delay was then enforced before the rakes were slid to within subjects' reach. This delay was intended to encourage consideration of the task and limit impulsive behaviour. The subjects were then able to select a rake by means of grasping the handle and pulling in the rake to retrieve the reward. A rake was considered as being selected once it had been pulled forward 10cm or more from the starting position. Only one selection was permitted per trial, with the unselected rake being removed from reach once a selection was made. Once subjects had either retrieved the reward (or not dependent on the rake selected), both rakes were pulled back out of subjects' reach, re-baited, and the next trial began. Maximum trial length was 30 minutes or until a rake selection had been made, with trials presented in blocks of ten. Inter-trial intervals of 20 seconds and inter-block intervals of five minutes were enforced in order to make trials and blocks distinct. If the subject failed to select a rake within the allocated trial time (30 minutes), testing was discontinued with this subject, continuing the following day. Subjects were required to achieve 8/10 correct selections of the functional rake on two consecutive blocks before they were considered as reaching criterion level, representing a highly significant level of selections of the functional rake. A maximum of two blocks (20 trials) a day were presented to an individual, with a maximum of either 60 or 120 trials presented to each subject. The different trial numbers were due to the best performing individuals receiving an additional six blocks to assess if they could reach criterion level (see section 3.3.1 for details). The position of the functional rake on either the left or right side of the testing platform was randomised, with the stipulation that it appeared on both sides equally and never more than twice on the same side successively during each block. Trials were recorded using a video camera positioned at the end of testing platform furthest from subjects for later behavioural coding. The position of the functional rake (on either the left or right side of the testing platform), and the rake selected was noted during testing.

3.2.3 Behavioural variables

As gibbons were not restrained during testing and were free to move around their enclosures, only time in proximity to the apparatus was considered as time available for learning. The area surrounding the testing platform ($1m^2$) was therefore designated as a target area, with behaviours only being coded whilst subjects were within this area. This distance was chosen given difficulties in determining if subjects were attending to the apparatus or other objects located close by when subjects were at a further distance. Total time within the target area was coded, along with the proportion of time in the target area attending to the apparatus (see Table 3.2). Subjects were classified as attending to the apparatus when they were visually oriented towards, or in physical contact with the apparatus. This included physical contact with any part of the apparatus through gibbons' hands, feet, or mouth. Time to first productive contact was coded as the time from when subjects entered the target area and attended to the apparatus, until the first rake contact that moved the rake towards the subject. The time and frequency of reward directed and non-reward directed contact with the rakes was coded. Reward directed contact was defined as any contact which moved the rake closer to the subject, with non-reward directed contact being coded if the subject spent 0.5 seconds or longer simply holding, or manipulating the rake in a manner that did not pull the rake towards them. This time limit was applied so the time taken to pick up the rake handle was not classified as non-reward directed contact.

The side that subjects approached the testing platform, and the side they were positioned on when a rake was selected, were also coded to assess whether poorly performing subjects were simply selecting the rake closest to them when they entered the target area. As subjects occasionally entered and exited the target area multiple times before making a selection, approach side was coded for the time subjects entered the target area and subsequently made a selection. The area was divided into 'left' and 'right' by means of the wooden divider placed centrally on the testing platform. Subjects were therefore classified as approaching and making a selection from either the left, right, or centrally (in line with the wooden divider). Behavioural coding was completed using 'Solomon Coder' (version beta 17.02.15), a free coding program available online (Péter, 2018). For analyses, age was categorised as adult (8+ years), sub-adult (6-8 years) and adolescent (3-5 years), as classified in Palombit (1994).

Table 3.2: Definitions of the behaviours coded from trial videos.

Behaviour	Description
Time in target area	Total time spent within the designated target area, both attending to, and not attending to, the apparatus.
Time in target area attending to apparatus	Visually attending to, or in physical contact with any part of the apparatus through hands, feet or mouth while in the target area.

Time to first productive rake contact	Time to first contact that moved the rake towards the subject in a reward directed manner. Taken from the time when subjects entered the target area and attended to the apparatus until first reward directed contact.
Reward directed contact	Duration and frequency of contacts with either rake which moved the rake closer to the subject.
Non-reward directed contact	Duration and frequency of contacts with either rake which did not move the rake closer to the subject. Contact was classified as non-reward directed if it occurred for 0.5 seconds or longer but did not move the rake closer to the subject.
Direction of approach	The direction of approach to the apparatus when the subject subsequently made a selection without leaving the target area again (left, centre, right).
Position on selection	The position of the subject in relation to the central line of the apparatus when a selection was made (left, centre, right).

3.3 Results

3.3.1 General performance overview

Of the 23 subjects who completed the minimum of 60 trials, only one (Win Bo) successfully reached criterion level performance, classified as correct selection of the functional rake 8/10 times over two consecutive blocks (see Table 3.3). Thirteen subjects displayed a side bias during the initial six blocks, choosing a rake positioned on the same side (left/right) of the testing platform across one or more blocks (see Table 3.3). After discounting individuals who were displaying a side bias, the six best performing individuals (taken as those who had selected the functional rake the highest number of times on the first 60 trials) were presented with a maximum additional six blocks to assess if further exposure to the task would result in improved performance. Lucia and Chan Thar both selected the functional rake 23/60 during the first six blocks, and were therefore the joint sixth best performing. However, Lucia was chosen for further testing as Chan Thar was housed with an individual undergoing medical treatment and it was preferential to keep disruption in this enclosure to a minimum.

Lucia and Truman reached criterion level performance within the additional six blocks. The other four subjects given further testing failed to reach criterion level, with two now displaying side biases across multiple blocks. Performance on the first two versus the last two blocks completed by each subject were evaluated using binomial tests to determine whether individuals were improving across blocks (see Table 3.3). Performance levels were improved in the last two blocks, when compared to the first two, for 11 subjects. However, the three subjects reaching criterion level performance were the only individuals performing above chance levels, during the last two blocks only. Performance levels for all individuals across all trials was also evaluated using binomial tests, with only Truman selecting the functional rake significantly above chance levels across all trials. Overall, findings suggested the majority of subjects failed to attend to the functionally relevant features of the rakes.

Nine subjects tested here had previously participated in raking in tasks (Cunningham, 2006; Cunningham et al., 2006, 2011). One of the nine individuals (Vok) failed to complete testing, losing interest in the task. None of the remaining eight subjects reached criterion level; six displayed side biases in the initial 60 trials and were not given the further testing consisting of an additional six blocks. Betty and Chloe did progress to further testing, and both displayed a criterion level performance on one or two separate blocks; however, neither reached criterion level. Previous experience was therefore assumed to not aid performance on the current task.

Table 3.3: Subjects' genus, age group and sex. Number of correct selections of the functional rake in each block completed by subjects (1 block = 10 trials). Blocks where subjects displayed a side bias that potentially hindered performance or reached criterion level performance are highlighted. Significance on binomial tests for number of correct selections of the functional rake across the total number of trials completed by each subject, the first two blocks (20 trials), and last two blocks are also indicated.

Subject	**Genus**	**Age Group**	**Sex**	**Number of correct trials per block**												**Number of correct trials**		
				1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total trials**	**First 2 blocks**	**Last 2 blocks**
Arthur	*Hoolock*	Adult	M	4	7	4	6	4	5							30/60	11	9
Betty	*Hoolock*	Adult	F	5	6	6	5	8	7	6	5	4	5	6	5	68/120	11	11
Chan Thar	*Hoolock*	Adult	F	8	5	5	4	4	6							32/60	13	10
Khin Maung Win	*Hoolock*	Sub-adult	M	4	3	5	6	6	5							29/60	7	11
Phy Gyi	*Hoolock*	Adult	F	6	5	5	4	5	5							30/60	11	10
U Maung Maung	*Hoolock*	Adult	M	7	6	6	4	5	5							33/60	13	10
U Myint Swe	*Hoolock*	Adolescent	M	6	4	5	6	3	5							29/60	10	8
Win Bo[1]	*Hoolock*	Adult	M	4	4	5	5	8	9							35/60	8	17**

Chloe	*Hylobates*	Adult	F	5	5	4	8	6	5	5	4	6	9	6	5	68/120	10	11
Domino	*Hylobates*	Adult	M	4	6	7	5	5	5							32/60	10	10
Iszie	*Hylobates*	Adolescent	F	3	5	4	5	5	5							27/60	8	10
Ivan	*Hylobates*	Adult	M	5	5	5	5	6	7							33/60	10	13
Khusus	*Hylobates*	Adult	F	4	4	5	5	7	5							30/60	8	12
Reg	*Hylobates*	Adult	M	5	5	5	4	5	5							29/60	10	10
Truman[1]	*Hylobates*	Adult	M	5	6	4	7	7	6	7	6	5	8	8		69/110**	11	16*
Tuk	*Hylobates*	Adult	F	6	3	6	5	5	7							32/60	9	12
Violet	*Hylobates*	Adolescent	F	5	8	3	4	6	5							31/60	13	11
Asteriks	*Nomascus*	Adult	F	6	3	5	3	5	5							27/60	9	10
Lucia[1]	*Nomascus*	Adolescent	F	3	6	6	4	5	8	6	4	8	8			58/100	9	16*
Paddy	*Nomascus*	Adult	M	4	3	5	6	4	6							28/60	7	10
Ricky	*Nomascus*	Adult	F	7	5	3	5	8	3							31/60	12	11
Karenina	*Symphalangus*	Adult	F	6	6	6	6	7	4	6	6	6	4	4	3	64/120	12	7
Marlow	*Symphalangus*	Adult	F	7	5	7	4	5	7	5	6	6	6	5	5	68/120	12	10

= Side bias

= Criterion level performance

* Significant at $p < 0.05$ level

** Significant at $p < 0.01$ level

[1] Reached criterion

The position of the functional rake was randomised, appearing equal times on the left and right side of the testing platform. Thus, in order to reach criterion level, subjects were required to select rakes positioned on both sides of the platform at equal rates. However, many subjects displayed side biases, repeatedly selecting a rake positioned on either the left or right side of the testing platform. The frequency of blocks displaying a side bias generally increased as testing progressed, suggesting subjects were reverting to this strategy when failing to learn to differentiate between the rakes. These findings are in line with Anderson (2012), who noted gibbons (*H. leucogenys*, n = 3, *S. syndactylus*, n = 2) developed a side bias when they appeared to become frustrated during an object permanence task, with one subject displaying a tendency to select the item closest to her.

Only five of the 23 subjects did not select the rake positioned on the same side of the testing platform (left/right) significantly across all trials (see Table 3.4). This five included the three subjects who reached criterion level, suggesting these subjects' selections were not guided by rake position as they learned to differentiate between the rakes and select the rake providing an attainable reward. Win Bo and Lucia did not display a preference for a particular side of the testing platform during the first 20 trials. Truman initially displayed a preference for rakes placed on the left side of the platform, significantly selecting from this side during the first 20 trials, but this preference was replaced by unbiased selections as testing progressed and he reached criterion level.

Of the 18 subjects who displayed a side bias across all trials, 12 individuals significantly approached the apparatus from either the left or right side and selected the rake positioned on the side of the platform they approached from (see Table 3.4). Thus, many subjects appeared to be simply selecting whichever rake was closest, suggesting they did not consider the task before making a selection. It is notable, however, that the strategy of always selecting the rake placed on the left or right resulted in reward retrieval rates of 50%. This strategy therefore ensures consistent reward retrieval in what may have appeared an uncertain choice situation, if gibbons were failing to discriminate between the rakes. The side chosen may be related to subject's lateral biases, or hand preference. Hand preference was assessed by measuring which hand subjects used most frequently to manipulate the rakes, with all subjects showing a significant individual bias (see Table 3.4). Relatively even numbers of individuals preferred to use their right (12 subjects) and left (11 subjects) hand, supporting recent findings of individual, but not population, biases in hand preference during a task requiring gibbons (*Hylobates*, *Nomascus* and *Symphalangus* sp.) to select objects by reaching through their enclosure mesh (Caspar, Mader, Pallasdies, Lindenmeier & Begall, 2018). Ten subjects approached from, and selected the rake positioned on, the side of the testing platform related to their preferred hand for manipulating the rakes (see Table 3.4). Manipulating the rakes required sufficient motor control, and subjects may therefore have situated themselves to allow for their preferred hand to draw in the rake.

Table 3.4: Subjects' preferred hand used for rake manipulations, the predominant side that subjects approached the testing platform from (left/right), and position of the rake selected on the platform (regardless of functional or non-functional). Data are collapsed across all trials, the first two blocks, and the last two blocks.

Subject	**Hand preference**	**Approach Side**	**Position of rake predominantly selected (significance on binomial test)**		
			Total trials	**First 2 blocks (20 trials)**	**Last 2 blocks (20 trials)**
Arthur	L**	n.s	R**	R**	R**
Betty	R**	n.s	n.s	R**	n.s
Chan Thar	R*	n.s	R**	R*	R**
Khin Maung Win	L**	L**	L**	L*	L**
Phy Gyi	L**	L**	L**	L**	L**
U Maung Maung	L**	n.s	R**	n.s	R**

U Myint Swe	L**	L**	L**	L**	L**
WinBo[1]	R*	n.s	n.s	n.s	n.s
Chloe	R**	R**	L**	n.s	L**
Domino	R**	n.s	L**	L**	L**
Iszie	R**	L**	L**	n.s	L**
Ivan	R**	R*	R**	R**	R*
Khusus	L**	R*	R**	n.s	R*
Reg	L**	L**	L**	L**	L**
Truman[1]	R**	R**	n.s	L**	n.s
Tuk	L**	L**	L**	n.s	L**
Violet	L**	L**	L**	n.s	L**
Asteriks	L**	n.s	n.s	n.s	n.s
Lucia[1]	L**	L**	n.s	n.s	n.s
Paddy	R**	n.s	R**	R*	n.s
Ricky	R**	R**	R**	R**	n.s
Karenina	R**	R**	R**	n.s	R**
Marlow	R**	R**	R**	n.s	R**

n.s = No significant preference

[1] Reached criterion

* Significant at $p < 0.05$ level

** Significant at $p < 0.01$ level

Generalized Estimating Equations (GEE), an extension of generalized linear models usable for repeated measures, were used to a build model to assess the effect of block number (1-6) on the number of correct responses to determine if performance was improving with experience. GEE is suited to analyses of small samples, with non-normally distributed, dependant data and can support many different types of variables such as binary or counts (Garson, 2013). The model was specified as Poisson loglinear with log link function, with number of correct selections of the functional rake as the dependent variable and genus, age group, sex, success (reached criterion level performance or not) and block number (1-6) as factors. Success was included as a factor to measure average frequency of functional rake selections per block, assessing if those subjects who reached criterion level performance consistently performed better than subjects failing to achieve criterion level. Only the first six blocks were considered as all subjects completed a minimum of 60 trials. The working correlation matrix was specified as 'independent' meaning trials were assumed to be independent of each other. For parameter estimates, *Symphalangus* was the reference category when evaluating genera differences. Block 6 was the reference category when evaluating differences across blocks. Age was categorised as adult (8+ years), sub-adult (6-8 years) and adolescent (3-5 years), with adolescent being the reference category.

Genus was a significant predictor of performance when age, sex, and success were held constant, with difference in performance levels between *Nomascus* and *Symphalangus* (GEE: $\beta = -0.18$, $\chi^2 = 30.98$, df = 1, $p < 0.01$) and between *Hylobates* and *Symphalangus* (GEE: $\beta = -0.09$, $\chi^2 = 19.53$, df = 1, $p < 0.01$), but not between *Hoolock* and *Symphalangus* GEE: $\beta = -0.07$, $\chi^2 = 3.39$, df = 1, $p = 0.07$). *Symphalangus* selected the functional rake the highest number of times, followed by *Hoolock* and then *Hylobates*, with *Nomascus* performing the poorest (see Figure 3.6). In general, performance levels did not differ greatly between the genera; *Symphalangus* selected the functional rake an average of 6/10 trials across the first six blocks, compared to the worst performing *Nomascus*, scoring an average of 5/10 correct selections.

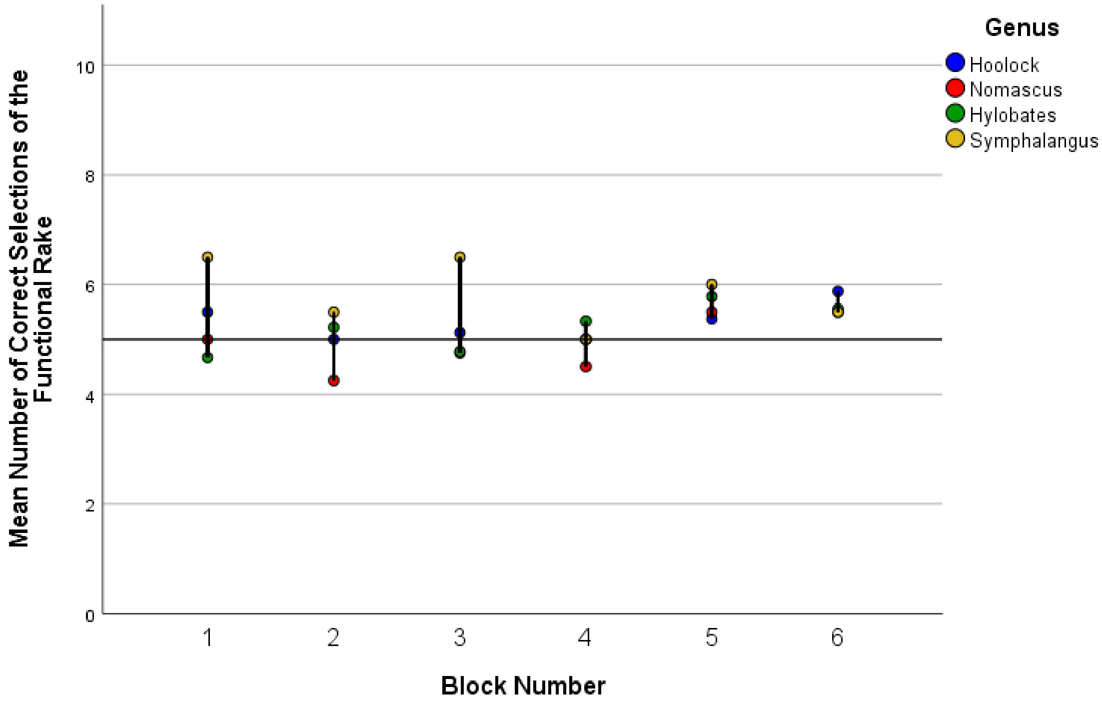


Figure 3.6: Mean number of correct selections of the functional rake across blocks 1 to 6 for each genus, with chance level performance marked.

Age also predicted performance (when all other factors were held constant) with significant differences between adults and adolescents (GEE: β = 0.08, χ^2 = 7, df = 1, $p < 0.01$), but not sub-adults and adolescents (GEE: β = -0.01, χ^2 = .032, df = 1, p = 0.86). In general adults performed better than either sub-adults or adolescents. No significant difference in number of correct selections of the functional rake was found between sexes (GEE: β = -0.03, χ^2 = 1.49, df = 1, p = 0.22), with females selecting the functional rake slightly more frequently than males. Unsurprisingly, success (reached criterion level or not) predicted performance with individuals that reached criterion level performance selecting the functional rake significantly more frequently, per block, than those who did not (GEE: β = 0.13, χ^2 = 46.4, df = 1, $p < 0.01$). There was no significant effect of block number on number of correct responses between blocks 1, 2, 4 or 5 and the reference block 6. There was however a significant difference in number of correct selections of the functional rake between blocks 3 and 6 (GEE: β = -0.11, χ^2 = 5.12, df = 1, p = 0.02). This is unlikely to be a meaningful difference as performance was generally poor with the mean number of correct selections being 5/10 across most blocks.

3.3.2 Evaluation of productive behaviours

3.3.2.1 Time spent in the target area

GEEs were used to build two models assessing the effect of genus, age, sex, success (reached criterion level performance), and block number (1-6) on total time spent in the target area and time spent in the target area attending to the apparatus (either visually inspecting or in physical contact with the apparatus). This assessed how long subjects spent engaged with task as testing progressed. Both models were specified as linear with identity link function, with either total time spent in the target area or time spent attending to the apparatus as the dependent variable. Genus, age group, sex, success and block number were set as factors. The working correlation matrix was specified as 'independent' meaning time values were assumed to be independent of each other. Reference categories were the same as in the previous GEE model (described in section 3.3.1).

Genus was a significant predictor across both models when all other factors were held constant. The difference in total time spent in the target area was significant between *Hoolock* and *Symphalangus* only (GEE: $\beta = 2.20$, $\chi^2 = 7.33$, df = 1, $p < 0.01$), with *Hoolock* spending the most time in the target area. For time spent in the target area attending to the apparatus, there were significant differences between *Hoolock* and *Symphalangus* (GEE: $\beta = 1.15$, $\chi^2 = 4.66$, df = 1, $p < 0.05$) who spent less time attending to the task, and between *Nomascus* and *Symphalangus* (GEE: $\beta = 1.2$, $\chi^2 = 12.46$, df = 1, $p < 0.01$), with *Symphalangus* again spending less time engaged with the task (see Figure 3.7). There was no significant difference between *Symphalangus* and *Hylobates*. Despite *Symphalangus* subjects selecting the functional rake most frequently, they spent the least time both in the target area in total, and attending to the apparatus, suggesting these individuals were processing the task more efficiently.

Sex was a significant predictor of time spent in the target area attending to the apparatus, but not total time in the target area. Females spent longer than males attending to the apparatus (GEE: $\beta = -1.52$, $\chi^2 = 7.73$, df = 1, $p < 0.01$). Neither age nor success significantly predicted the amount of time spent in the target area in total, or time attending to the apparatus. However, those subjects that reached criterion level generally spent less time in the target area in total and less time attending to the apparatus, again suggesting successful subjects were able to process the task demands more efficiently. Despite spending the longest time in the target area, sub-adults spent the least time attending to the apparatus.

Block number was a significant predictor for both total time in the target area and time spent attending to the apparatus when all other factors were held constant. For total time in the target area, the difference between the reference block 6 and block 1 (GEE: $\beta = 5.73$, $\chi^2 = 16.94$, df = 1, $p < 0.01$), and between block 6 and block 2 (GEE: $\beta = 3.44$, $\chi^2 = 7.94$, df = 1, $p < 0.01$) were significant. For time spent attending to the apparatus, the differences between block 6 and block 1 (GEE: $\beta = 3.08$, $\chi^2 = 21.54$, df = 1, $p < 0.01$), block 2 (GEE: $\beta = 2.06$, $\chi^2 = 20.17$, df = 1, $p < 0.01$), and block 3 (GEE: $\beta = 0.96$, $\chi^2 = 5.94$, df = 1, $p < 0.05$) were significant. Subjects spent less time in the target area and attending to the apparatus as blocks progressed (Table 3.7). As performance levels did not increased across blocks 1-6, this may represent more impulsive selections and decreasing interest in the task as testing progressed.

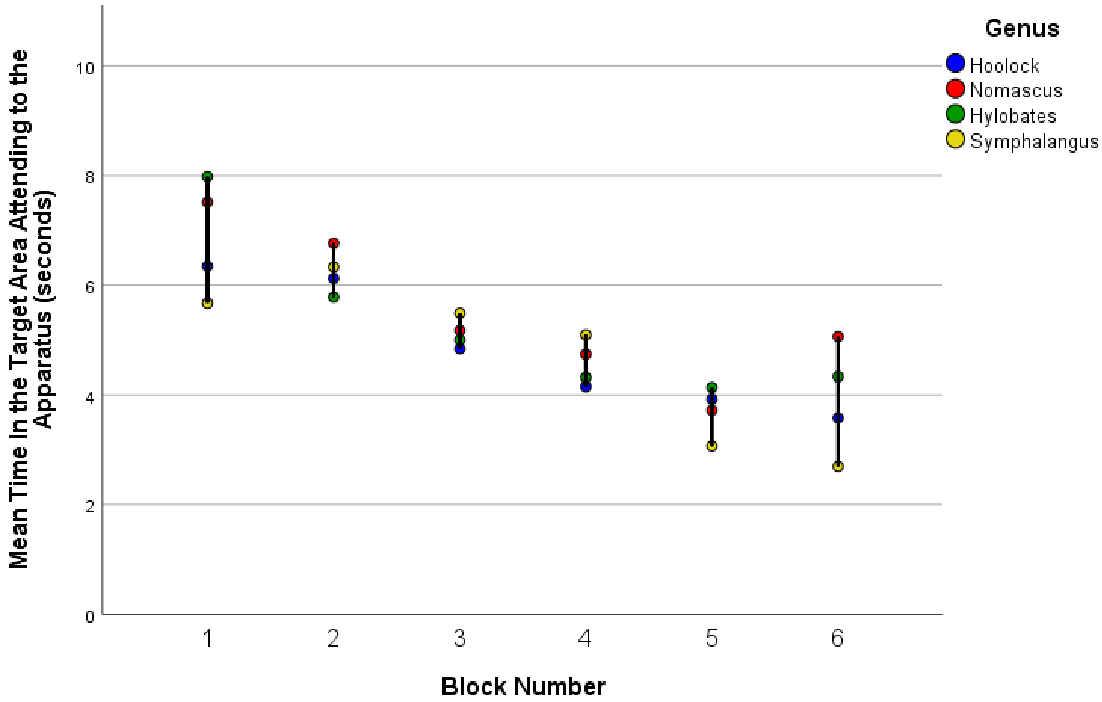


Figure 3.7: Mean time spent in the target area attending to the apparatus for each genus across blocks 1 to 6.

A further model was built using GEEs to assess the effect of genus, age, sex, success (reached criterion level or not), and block number on time to first productive contact with the rakes, assessing cognitive processing time before selections were made. The model was identical to those built to assess time in the target area except time to first productive contact was the dependent variable. Genus was a significant predictor of time to first productive rake contact when all other factors were held constant (see Figure 3.6). *Symphalangus* were the fastest to first productive contact, and *Hoolock* the slowest. These findings again suggest that *Symphalangus* subjects were more task focussed with higher cognitive processing speeds, something which may have led to the better performance levels shown by these subjects. The differences between *Nomascus* and *Symphalangus* (GEE: $\beta = 0.96$, $\chi^2 = 10.76$, df = 1, $p < 0.01$), and *Hoolock* and *Symphalangus* (GEE: $\beta = 0.98$, $\chi^2 = 3.94$, df = 1, $p < 0.05$) were significant. Likewise, although success did not significantly predict time to first productive contact, subjects that reached criterion level were generally faster to first productive contact, suggesting as subjects learned to differentiate between the rakes, less time was spent considering the task before making a selection. When evaluating age and sex differences, the better performing female and adult groups were slower to first productive contact than males and the younger subjects, suggesting less impulsive behaviour in these groups. Age was not a significant predictor; sub-adults, who selected the functional rake at the same level as adolescents, were however, faster to first productive rake contact, with adults taking longer. The difference between males and the slower female group was significant (GEE: $\beta = -1.34$, $\chi^2 = 6.87$, df = 1, p = < 0.01).

Subjects took longer to first productive contact during block 1, with the time taken decreasing across blocks 1-5. Time to first productive contact increased again slightly in block 6, but the general trend for faster time to first contact remained. These patterns were the same as found for time in the target area, again suggesting subjects were making more impulsive actions as testing progressed. Block number significantly predicted performance when all other factors were held constant, with the significant differences between the reference block 6 and block 1 (GEE: $\beta = 2.31$, $\chi^2 = 19.99$, df = 1, p = < 0.01), block 2 (GEE: $\beta = 1.76$, $\chi^2 = 15.01$, df = 1, p = < 0.01), and block 3 (GEE: $\beta = 0.84$, $\chi^2 = 4.48$, df = 1, p = < 0.05).

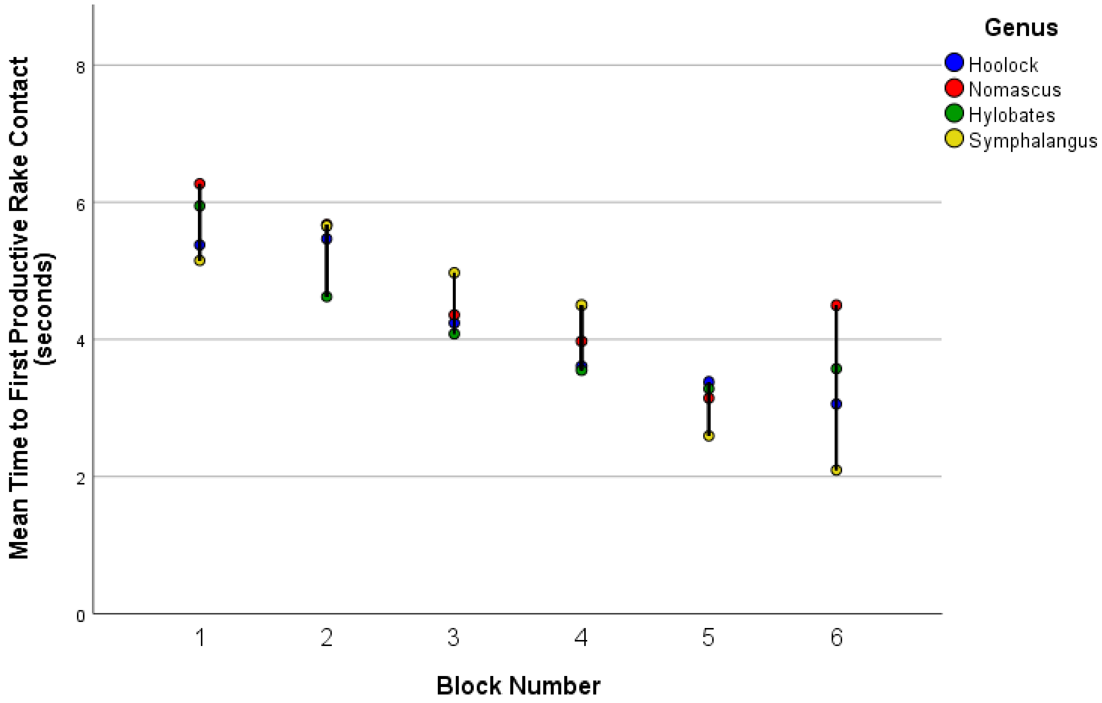


Figure 3.8: Mean time to first productive contact with the rakes for each genus across blocks 1 to 6.

3.3.2.2 Rake contacts

Two models were built using GEEs to assess the effect of genus, age, sex, success (reached criterion level performance or not), and block number (1-6) on time spent in contact with the rakes, assessing if more effective rake manipulations were being made as testing progressed. The model specifications were as previously described (see section 3.3.1 for details), however, time spent in reward directed contact, and time spent in non-reward directed contact were now set as the dependent variables. Time spent in non-reward directed contact was minimal across all trials, with the model assessing time spent in non-reward directed contact producing no significant effects. As such, this model is excluded from further analyses.

Genus was a significant predictor of reward directed contact when all other factors were held constant (see Figure 3.9). *Symphalangus* spent the least time in reward directed rake contact, with *Hylobates* spending the most time. The difference between *Symphalangus* and both *Nomascus* (GEE: $\beta = 0.18$, $\chi^2 = 7.95$, df = 1, $p < 0.01$) and *Hylobates* (GEE: $\beta = 0.34$, $\chi^2 = 18.49$, df = 1, $p < 0.01$) was significant. Although *Hylobates* spent the most time manipulating the rakes, this did not result in improved performance as this was the second worst performing genus. Again, *Symphalangus* appear to have exhibited more focussed behaviour than the other genera, achieving the highest number of correct selections of the functional rake despite spending the least time in reward directed rake manipulations.

Age was a significant predictor of reward directed contact (adults and adolescents GEE: $\beta = -0.23$, $\chi^2 = 7.6$, df = 1, $p < 0.01$; sub-adults and adolescents GEE: $\beta = -0.39$, $\chi^2 = 13.89$, df = 1, $p < 0.01$). Despite spending the most time in the target area, sub-adults spent the least time in rake manipulations and attending to the apparatus, suggesting they engaged with the task less than other age groups. Adolescents spent the most time in reward directed contact and attending to the apparatus. However, this increased time did not result in an improved performance, as adolescents performed similarly to sub-adults. Females, who selected the functional rake more frequently than males, spent longer in reward directed contact (GEE: $\beta = -0.21$, $\chi^2 = 8.79$, df = 1, $p < 0.01$). Despite being slower to first make contact with the rakes, females spent longer engaging with the task, both through rake manipulations and time spent in the target area.

Success was not a significant predictor; however, successful subjects, who reached criterion level, spent longer in reward directed contact, suggesting these subjects were making more efficient manipulations once they had learned to differentiate between the rakes. The time subjects spent in reward directed contact with the rakes decreased as trials progressed from block 1 to 6. Given that performance levels did not increase across the blocks and that time spent in the target area and attending to the task decreased, this may again represent declining interest in the task, rather than an increase in efficient manipulations. Block number was a significant predictor of reward directed contact when comparing block 6 to block 1 (GEE: $\beta = 0.3$, $\chi^2 = 13.84$, df = 1, $p < 0.01$), and block 2 (GEE: $\beta = 0.17$, $\chi^2 = 11.64$, df = 1, $p < 0.01$) (Figure 3.9).

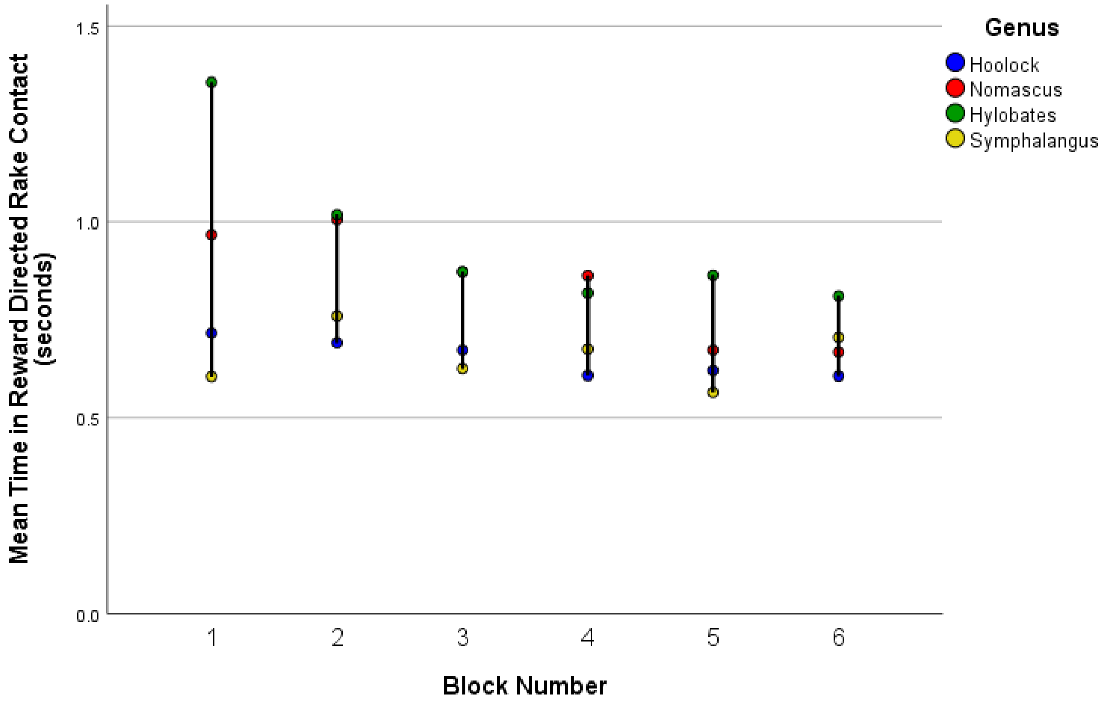


Figure 3.9: Mean time spent in reward directed contact with the rakes for each genus across blocks 1 to 6.

3.3.3 Order effects

In some cases, two or more subjects from the same enclosure were tested. Although individuals were separated for testing (with the exception of enclosure 5) and the positioning of the apparatus aimed to limit visual access for the not-being-tested individual, there was a possibility that being in close proximity to their partner during testing may have provided some advantage in the learning process. Therefore, order effects were assessed using Mann-Whitney U tests for all enclosures where two or more individuals participated in testing. Time to first productive contact with the rakes was compared over the first ten trials. No significant difference was found for Chan Thar and Win Bo ($U = 37$, $z = -0.98$, $p = 0.35$), Violet and Truman ($U = 49$, $z = -0.08$, $p = 0.97$), Phy Gyi and Arthur ($U = 33$, $z = -1.29$, $p = 0.22$), Marlow and Karenina ($U = 41$, $z = -0.68$, $p = 0.53$), or Asteriks and Lucia ($U = 39$, $z = -0.83$, $p = 0.44$).

Significant differences were found between Chloe and Ivan in enclosure 1 ($U = 7.5$, $z = -3.22$, $p < 0.01$), and Ricky and Paddy in enclosure 13 ($U = 21$, $z = -2.19$, $p < 0.05$). Ivan, who was tested second, took longer to first productive contact than Chloe, with both subjects selecting the functional rake on 5/10 trials, suggesting testing order did not affect performance. Paddy, the second individual tested, was faster to first productive contact than Ricky; however, he selected the functional rake on fewer trials, suggesting performance was not aided by testing order. Ricky was likely slower than Paddy due to her movement being slightly hindered following the removal of one of her legs during a medical procedure.

There were two enclosures were three individuals were tested. For enclosure 10, subjects were tested separately. The first individual tested, Betty, took longest to first productive contact (mean = 7.79 seconds), with the second tested, U Myint Swe, taking an average of 5.7 seconds, and finally Khin Maung Win, the third tested, taking an average of 3 seconds. There was no significant difference in time to first productive contact between Betty and U Myint Swe ($U = 28$, $z = -1.67$, $p = 0.11$). There was a significant difference between Betty and Khin Maung Win, ($U = 13$, $z = -2.8$, $p < 0.01$), and between U Myint Swe and Khin Maung Win ($U = 18.5$, $z = -2.39$, $p < 0.05$). Considering performance levels, Betty selected the functional rake on 5/10 trials, with U Myint Swe selecting the functional rake on 6/10 trials and Khin Maung Win on 4/10 trials. Thus although the order of testing may have led Khin Maung Win to make his first productive contact with the apparatus faster, performance levels did not appear to be greatly improved.

Individuals were tested in a group setting in enclosure 5 containing three subjects as they could not be separated. Tuk was the first to be tested and the fastest to first productive contact (mean = 4.65 seconds), followed by Iszie who took the longest to first productive contact (mean = 13.39 seconds), with Domino tested last (mean = 5.09 seconds). There was no significant difference between Tuk and Domino ($U = 41$, $z = -0.68$, $p = 0.53$) or Iszie and Domino ($U = 28$, $z = -1.66$, $p = 0.11$), however, there was a significant difference between Tuk and Iszie ($U = 22.5$, $z = -2.08$, $p < 0.05$). Tuk selected the functional rake most frequently (6/10 trials), followed by Domino (4/10 trials), and Iszie (3/10 trials). Testing order therefore did not appear to aid performance, as both Iszie and Domino were slower and performed poorer than Tuk.

3.4 Discussion

The current research aimed to assess if gibbons could attend to the functionality of two different rakes during a zero-order raking in task. Performance was poor with only three subjects reaching criterion level performance despite extended experience. Successful individuals spent less time attending to the apparatus and were faster to first productive rake contact than unsuccessful subjects suggesting they were making more focused manipulations, potentially due to faster processing of the task demands. These subjects also spent longer in reward directed contact and less time in non-reward directed rake contacts, suggesting they were being more efficient as they learned to differentiate between the rakes. It is also of note that the three successful individuals did not display a side bias, although it is not clear from the present data whether this was as a result of a clearer understanding of the task requirements, or conversely the lack of side bias led to better performance.

Genera differences also supported the proposal that better performing individuals spent less time engaged with the task; however, the pattern of results across genera was inconsistent. *Symphalangus*, the genus who selected the functional rake most frequently, spent the least time attending to the apparatus and in rake contact, were also the fastest to first productive interactions with the apparatus. *Symphalangus* therefore appeared to be making more efficient use of their time engaged with the task than the other genera, potentially due to a better understanding of the task; however, with only two *Symphalangus* subjects included, and the overall poor performance of most gibbons, conclusions are tentative. The worst performing genus, *Nomascus*, spent longer in the target area compared to *Symphalangus*, attending to the apparatus for longer, being slower to first productive rake contact and spending more time in reward directed contact. These preliminary results imply that those subjects that understood the task reacted more quickly and were more efficient in their actions with the apparatus.

Females were slower to obtain the reward than males, taking more time before first productive contact with the rake, and spending more time spent in rake contact, suggesting a more cautious approach to the task. This is consistent with Cunningham et al. (2011) who conducted an assessment of the effect of prior experience on problem solving in gibbons (*Hoolock*, *Nomascus*, *Hylobates* and *Symphalangus* species) during a raking in task. Females without prior exposure to the raking in apparatus were slower to task solution than males with no prior exposure. Cunningham et al. (2011) suggest females may be more cautious of novel objects as the have more to lose in fitness terms, especially if they have dependent offspring. In contrast to the current research, Cunningham et al. (2011) also found females, both with and without prior experience of the apparatus, were faster to first contact with the apparatus than males, though not necessarily in a productive way. Six of the 13 females in the current study had dependent offspring (a greater number than in Cunningham et al. 2011), and a number of these individuals were tested in their family groups. This may have resulted in the increased caution displayed before first contacting the rakes by the current female subjects.

The pattern of behaviour seen between the age groups was less clear; however, adolescents tended to spend longer engaging with the task than either sub-adults or adults. This supports previous findings from Cunningham (2006), who found time spent in the target area generally decreased as age increased during a raking in task. Conducting a review of innovation and problem solving in animals, Griffin and Guez (2014) concluded juveniles appeared to be less neophobic than adults; however this does not reliably lead to improved problem solving skills. This conclusion is supported by research from Kendall, Coe and Laland (2005) who found that attentiveness to a novel extractive foraging task decreased as age increased in callitrichid monkeys. However, adults were faster to successfully manipulate the task than non-adults, suggesting more task-focussed, innovative behaviour in older individuals. Innovation may build upon other skills, with the increased object manipulation competency and experience of adults leading to a greater level of innovative behaviour than that seen in younger individuals (Reader & Laland, 2001). The current findings support these proposals, with adolescent subjects spending longer engaged with the task but displaying less successful behaviour than adult individuals.

Spaulding and Hauser (2005) proposed primates may possess an innate ability to attend to functionally relevant features of objects, with this mechanism requiring task-relevant experience in order for individuals to reach task sufficient performance. Tamarins (*S. oedipus*), capuchins (*C. apella*), marmosets (*C. jacchus*) and vervet monkeys (*C. aethiops*) have displayed the ability to attend to functionally relevant and irrelevant features of objects during a raking in task (Cummins-Sebree & Fragaszy, 2005; Fujita et al., 2003; Hauser, 1997; Hauser et al., 2002; Santos et al., 2006; Spaulding & Hauser, 2005). It was expected that gibbons would learn to attend to the functionally relevant perceptual features of rakes over an extended testing period, given gibbons' phylogenetic position and comparable performance to chimpanzees (*P. troglodytes*) on object manipulation tasks (Beck, 1967; Cunningham, 2006; Cunningham, et al., 2006; Inoue & Inoue, 2002).

The current findings do not support this hypothesis, as the gibbons failed to attend to differences between the two rakes over extended testing. The majority of subjects performed near chance levels, with only three individuals reaching criterion level performance and just one individual performing significantly above chance levels across all trials. Subjects generally spent longer engaged with the task during the first two blocks, likely due to the novelty of the task, with time spent engaged generally decreasing as testing progressed, suggesting lack of understanding (and so success), may have decreased motivation in later trials. The notion that gibbons did not attend to the functional features of the tools is further supported by the development of a side bias in multiple individuals. Even subjects who did not initially demonstrate side bias, and performed above chance in some blocks, later developed a side bias. This suggests subjects reverted to this strategy, which ensured reward retrieval on 50% of trials, when failing to discriminate between the rakes.

The task used in the current research was a partial replication of the 'inverted rake problem' developed by Povinelli and Reaux (2000). Giving a limited number of trials, Povinelli and Reaux (2000) and Santos et al. (2006) assessed spontaneous understanding of functionality during this task, finding subjects failed to significantly select the functional rake. Chimpanzees (*P. troglodytes*) selected the correct rake on 57% of trials, with tamarins (*S. oedipus*) selecting correctly on 52% of trials, and vervet monkeys (*C. aethiops*) on 60% of trials. Gibbons were provided with a greater number of trials, assessing if task relevant experience would aid subjects in learning to attend to functionally relevant features of the rakes. Considering the three individuals who reached criterion level performance in the current research, Win Bo, who reached criterion within the initial 60 trials, and Lucia who required 100 trials to reach criterion, both selected the functional rake on 58% of trials. Truman, who performed best over all trials, selected the functional rake on 63% of 110 trials. Despite the extended testing period of the current research, providing more opportunity to learn about the functional properties of the rakes, the successful gibbons performed at comparable levels to those species assessed with minimal exposure to the task.

It is possible the rakes in the current task were too similar in appearance, especially when viewed from above (the angle which subjects most often approached the apparatus) with this similarity obscuring differences in functionality and impairing performance. That is, by presenting rakes appearing visually similar (especially when approaching from above), it was not clear to the apes what dimension they should be attending to, leading to the low number of subjects reaching criterion. The series of experiments conducted in Chapter 4 therefore introduced a more salient difference in appearance between the rakes, aiming to reduce the cognitive demands of the task by making the ‘difference’ more obvious. Furthermore, the task used in the current research required consideration of the three dimensional relationship between the rakes and reward. This may present a more difficult task than those where the rewards position in relation to the rake can be taken in to account when discriminating between functional and non-functional objects. Chapter 4 therefore also presented several experiments with a food reward either positioned in-front of rakes or not, assessing if gibbons displayed a comparable performance to New World monkeys on this easier task.

Chapter 4: Attending to functional features of rakes

4.1 Introduction

It was hypothesised that the similarity in appearance of the rakes during Chapter 3 may have obscured differences in functionality and resulted in gibbons' (*Hylobatidae*) poor performance. Subjects frequently approached the testing platform from above, from which angle the rakes may have looked similar in shape, with differences in functionality not being particularly salient. Furthermore, the task used during pilot testing required understanding of the three dimensional relationship between the rakes and food reward, considering whether the rake would come in to contact with the food reward or not. This may present a more difficult challenge than tasks where consideration of the rewards location in relation to the rakes, either positioned in front of the rake or not, can be taken in to account. Povinelli and Reaux (2000) suggested naïve chimpanzees (*Pan troglodytes*) behaviour may be explainable by consideration of whether the reward is 'contained' in front of a tool or not when first attempting an inverted rake task with similar apparatus to that used in pilot testing (see Chapter 2, section 2.2 for further details). The increased difficulty of tasks requiring understanding of physical contact between rakes and rewards is further supported by findings of Santos, Pearson, Spaepen, Tsao and Hauser (2006, see Chapter 2, section 2.2 for further details) who found tamarins (*Saguinus oedipus*) and vervet monkeys (*Cercopithecus aethiops*) performed poorer on an inverted rake task than a task requiring consideration of whether a reward was placed inside or outside a curved hook at the end of a cane. Thus, the increased difficulty of the task presented in pilot testing, in combination with perceptual differences between the rakes being obscured, may have resulted in gibbons' poorer performance when compared to species of New World monkeys which have been found to attend to functionally relevant features of tools during a raking in task (Cummins-Sebree & Fragaszy, 2005; Fujita, Kuroshima & Asai, 2003; Hauser, 1997; Hauser, Pearson & Seelig, 2002; Spaulding & Hauser, 2005).

In order to address these issues, a series of four experiments were conducted, all of which presented one functional and one non-functional rake during a raking in task similar to that given during pilot testing (Chapter 3; see Table 4.2 and Figures 4.1-4.4 for rake descriptions). A more obvious difference in appearance between the rakes was introduced in order to reduce the cognitive demands of the task by making the ‘difference’ more obvious. During Experiments 1, 2 and 4, the reward was positioned in front of the functional rake, but to the side of the non-functional rake, and therefore not ‘contained’ in front of it. Experiment 3 presented a repeat of the apparatus used during pilot testing, requiring consideration of physical contact between the rakes and reward. This allowed for investigation of which aspects of the task subjects were attending to when making selections and assessment of subject’s ability to transfer knowledge of functionally relevant features across different rake sets.

Experiment 1 presented the functional rake used during pilot testing (see section 3.2.2) with a new cross shaped non-functional rake, each with a distinctive colour. During this experiment, subjects may therefore be learning about the functionality of the two rakes by attending to differences in colour, rake shape, or functionality. New World monkeys attended to perceptual features during a raking in task, differentiating between rakes with features that might affect functionality (e.g., shape, size) and those with features that did not (e.g., colour, texture, see Chapter 2, section 2.2) (Cummins-Sebree & Fragaszy, 2005; Fujita et al., 2003; Hauser, 1997; Hauser et al., 2002; Spaulding & Hauser, 2005). It was expected gibbons would also display the ability to differentiate between the two rakes, given the more obvious differences between them (see Figures 4.1-4.4).

Fujita et al. (2003) found capuchins (*Cebus apella*) continued to preferentially select a functional rake during a raking in task when presented with a rake set in a different colour, or a different shape from that used during training, suggesting the ability to generalise across different tool sets rather than learning being tied to specific perceptual features (see Chapter 2, section 2.2 for further details). Subjects who reliably selected the functional rake during Experiment 1 were presented with a series of experiments manipulating different aspects of the rakes, assessing if gibbons were also capable of selecting rakes based on relevant features, ignoring properties that would not affect usefulness. Success in one experiment was required before progression to the following experiment. Experiment 2 presented the same rake set as Experiment 1; however, the colour difference was now removed. This allowed for assessment of whether subjects' performance during Experiment 1 was guided by colour differences between the rakes, and encouraged subjects to attend to either the rake shape or functionality. Experiments 3 and 4 assessed if gibbons were capable of transferring their learning across different rake shapes, once they reliably selected the functional rake during Experiments 1 and 2, to disentangle whether shape or function was being used to guide behaviour.

Experiment 3 presented the rake set used during pilot testing (see Chapter 3, section 3.2.2). Thus, the functional rake was the same rake used throughout Experiments 1 and 2, with the non-functional rake changing from the cross shaped rake to a rectangular non-functional shape. This task therefore required consideration of whether the rakes would come in to contact with the reward, which was placed in front of both the functional and non-functional rake. Povinelli and Reaux (2000) found chimpanzees (*P. trogylodyte*s) performed around chance levels during a similar task (see Chapter 2, section 2.2 for further details), selecting a functional rake on 57% of trials. Likewise, Santos et al. (2006) found vervet monkeys (*Cercopithecus aethiops*) selected a functional rake on 60% of trials, and tamarins (*Saguinus oedipus*) 52% of trials during a replication of Povinelli and Reaux's (2000) inverted rake task. However, these studies assessed subjects' spontaneous understanding of the inverted rake task, with each subject given a limited number of trials. In contrast, gibbons failed this task even when given an extended number of trials to learn to attend to differences in functionality between rakes (see Chapter 3, section 3.3). Experiment 3 therefore assessed if these subjects could now display task sufficient performance following learning which aspects of the rakes were relevant during Experiments 1 and 2, transferring their knowledge to this rake configuration.

Experiment 4 again presented the cross shaped non-functional rake, with a new functional rake of a different shape to that used across Experiments 1-3. As subjects had been presented with the same functional rake across previous experiments, this experiment assessed whether they could transfer knowledge of functionality to this new rake. The number of trials given during Experiments

3 and 4 was limited, assessing subjects' initial response to the changing shape of the rakes rather than allowing prolonged opportunities for learning. If subjects were basing their selections during Experiments 1 and 2 solely on rake-end shape, it was expected performance levels during these experiments would initially be poorer as subjects re-learned the correct response. If gibbons were capable of attending to the functionality of different rakes, with learning not tied to specific perceptual features, it was expected they would transfer this knowledge between rake sets, with frequency of functional rake selections remaining high during the initial trials of each experiment.

Success across all four experiments would suggest that, like New World monkeys, gibbons were capable of differentiating between functional and non-functional rakes, attending to perceptual features which both did, and did not impact functionality (Cummins-Sebree & Fragaszy, 2005; Fujita et al., 2003; Hauser, 1997; Hauser et al., 2002; Spaulding & Hauser, 2005). As noted previously, the extended time available for learning across testing with the current subjects does not give a comparable performance to the assessments of spontaneous understanding during inverted rake tasks with chimpanzees, vervet monkeys and tamarins (Povinelli & Reaux, 2000; Santos et al., 2006). However, given that the rewards position in relation to the rake could not be used to guide selections during Experiment 3, success during this experiment would suggest gibbons can transfer knowledge of functionality to a rake set which also required consideration of the physical contact between the rake and reward.

4.2 General methods

4.2.1 Subjects

Subjects were 20 gibbons housed at the Gibbon Conservation Center (see section 3.2.1 for details of housing), which had participated in pilot testing (described in section 3.2.2). This included seven eastern hoolock (*Hoolock leuconedys*), four javan (*Hylobates moloch*), three northern white cheeked

(*Nomascus leucogenys*), five pileated (*Hylobates pileatus*) and one siamang gibbon (*Symphalangus syndactylus*) (see Table 4.1). Of the five remaining subjects who completed trials during pilot testing but did not continue with these experiments, three (Vok, Hmawe Ni and U Maung Maung) became disinterested with the task and so did not complete the necessary trials, one (Paddy) was relocated to another facility after pilot testing was complete, and the final subject (Karenina) passed away.

Subjects were separated and tested individually with the exception of enclosures 2, 5, 11 and 13. Although separation would have been preferable, the decision to test these individuals whilst in their groups took into account welfare issues, with these individuals becoming distressed when separated. Khusus was the only individual from enclosure 2 to participate; however, her two offspring were present during testing. In enclosure 11, Phy Gyi and Arthur were separated for testing; however, there were three offspring in this enclosure, with the youngest staying with Phy Gyi during testing and the older two infants remaining with Arthur. Ricky was the only individual who participated in enclosure 13, but she remained in her family group during testing, which included one adult male and three offspring. Subjects in enclosure 5 were tested in their family group, which included three individuals participating in testing and one additional infant. Within this enclosure the apparatus remained out of reach until the focal individual approached the testing platform. Subjects therefore did not receive additional trials during testing but may have witnessed other subjects in their enclosure participating in the task. Order effects were analysed for all enclosures where two or more subjects participated (see results sections for each experiment for analyses).

Table 4.1: Subjects' species, age, sex, housing information, details of which experiments they participated in and whether they reached criterion level performance or not.

Subject	Enclosure	Species	Age	Sex	Group	Exp. 1	Exp. 2	Exp. 3	Exp. 4
Chan Thar	6	*H. leuconedys*	9	F	Male/female pair	✓	✓	✓	✓
Win Bo[1]	6	*H. leuconedys*	12	M	Male/female pair	✓	✓	✓	✓
Betty	10	*H. leuconedys*	16	F	Female & male/male siblings	✓	✓	✓	✓
Khin Maung Win	10	*H. leuconedys*	7	M	Female & male/male siblings	✓	✓	✓	X
U Myint Swe	10	*H. leuconedys*	6	M	Female & male/male siblings	✓	✓	✓	✓
Arthur*	11	*H. leuconedys*	19	M	Family	✓	✓	X	
Phy Gyi*	11	*H. leuconedys*	12	F	Family	✓	✓	X	
Chloe	1	*H. moloch*	25	F	Family	✓	✓	X	
Ivan	1	*H. moloch*	41	M	Family	✓	X		
Khusus*	2	*H. moloch*	20	F	Family	✓	X		
Reg	4	*H. moloch*	15	M	Solitary	✓	✓	X	
Domino*	5	*H. pileatus*	20	M	Family	✓	X		
Iszie*	5	*H. pileatus*	4	F	Family	✓	✓	✓	✓
Tuk*	5	*H. pileatus*	22	F	Family	✓	✓	X	
Violet	9	*H. pileatus*	6	F	Solitary	✓	✓	✓	✓
Ricky*	13	*N. leucogenys*	30	F	Family	✓	✓	✓	X

Asteriks	15	*N. leucogenys*	15	F	Family	✓	✓	✓	✓
Marlow	14	*S. syndactylus*	10	F	Solitary	✓	✓	✓	✓

✓ = Reached criterion level

X = Did not reach criterion level

* Not separated for testing

[1] This subject initially completed a repeat of the pilot testing task (described in section 3.2.2) before moving to this series of experiments (see section 4.8).

4.2.2 General experimental procedure and apparatus

The general procedure was consistent across all four experiments and similar to pilot testing (see Chapter 3, section 3.2.2), with only rake type changing. Subjects were required to select between two rakes, one functional and one non-functional, presented simultaneously on a testing platform outside their enclosure. Following baiting of both rakes with a small food reward, a delay of five seconds was enforced before the rakes were pushed to within subjects' reach. This delay was intended to limit impulsive behaviour, encouraging attention to the task prior to making a selection. Gibbons were considered to have made a selection if they drew a rake towards them by at least 10cm. The un-selected rake was then removed from reach, ensuring only one selection was made per trial.

Trials were presented in blocks of ten, with subjects receiving a maximum of 20 trials in a day. A maximum of 12 blocks (120 trials) was given during Experiments 1 and 2, and a maximum of three blocks (30 trials) during Experiments 3 and 4. Trial length was a maximum of 30 minutes, with inter-trial intervals of 20 seconds and inter-block intervals a minimum of five minutes. If no selection was made within the allocated trial time, testing was discontinued with this subject, continuing the following day. Criterion level was set as 8/10 correct selections of the functional rake on two consecutive blocks. Once subjects achieved this, they proceeded to the next experiment. Any subject who did not reach criterion was excluded from further testing.

The position of the functional rake was randomised, appearing on the left and right side of the testing platform in equal numbers, but never on the same side more than twice in succession within a block. The food rewards used throughout were blueberries, a preferred food not normally included in the gibbon's diet. Testing was recorded with a video camera for later behavioural coding, with a note of functional rake position (left/right) and rake selected (functional/non-functional) taken during testing.

Rakes were presented on a raised platform (as described in section 3.2.2) situated outside the subjects' enclosure. Rake handles consisted of a polypropylene rod (as described in section 3.2.2), with a variety of functional and non-functional rake ends used (see Figures 4.1-4.4), as described in Table 4.2. In summary, during Experiment 1, one rectangular functional (F1) and one cross shaped non-functional rake (NF2) were presented with either ‘red’ or ‘green and yellow’ striped tape covering the rake handles, differentiating them on shape, functionality and colour. The same rakes were presented in Experiment 2, however, the colour difference was removed and both rake handles were black. Experiment 3 again presented the functional rectangular rake (F1), now with a rectangular non-functional (NF1) rake, both coloured black. Finally, Experiment 4 presented a new smaller rectangular functional rake (F2) and the previously presented cross shaped non-functional (NF2) rake, again both coloured black (further details and justifications for rake changes included in Table 4.2 and experiment specific sections; 4.3.1, 4.4.1, 4.5.1 and 4.6.1).

Figure 4.1: F1 = Rectangular functional rake and NF2 = Cross shaped non-functional rake used in Experiments 1 and 2.

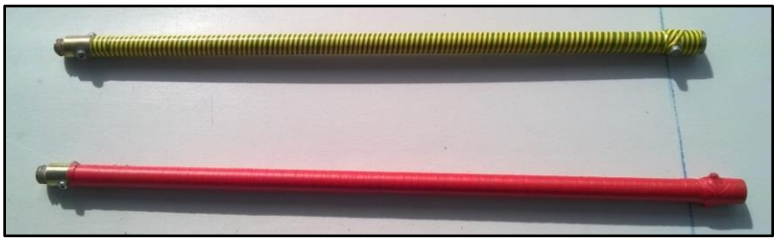


Figure 4.2: Coloured rake handles used in Experiment 1.

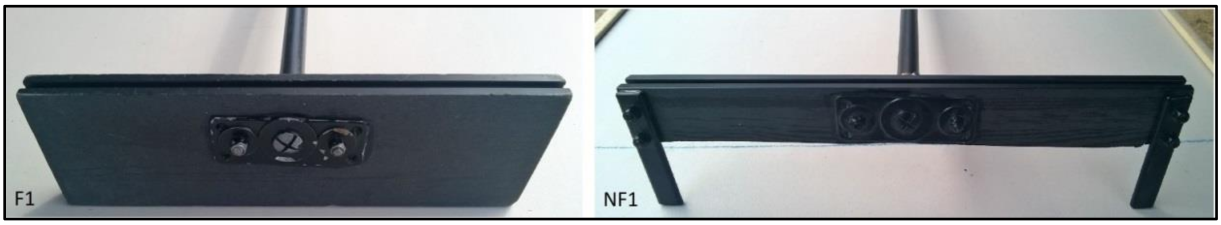


Figure 4.3: F1 = Rectangular functional rake and NF1 = Rectangular non-functional rake, used in Experiment 3.

Figure 4.4: F2 = Thin rectangular functional rake and NF2 = Cross shaped non-functional rake used in Experiment 4.

Table 4.2: Details of functional and non-functional rake end shapes used throughout Experiments 1-4.

Rake end type	**Description**
F1: Functional rectangular	34cm by 10.5cm rectangle. Previously presented during pilot testing (see Chapter 3), usable to draw a food reward within reach if selected.
F2: Thin functional rectangular	34cm by 5cm rectangle. Presented a visually different, but still functional, rake useable to draw a food reward within reach.
NF1: Rectangular non-functional	34cm by 10.5cm, with a smaller rectangle cut out the middle of the bottom (30cm by 6cm), leaving two legs of 2cm on either side.

	Previously presented during pilot testing (see Chapter 3), non-functional for reward retrieval as the rake passes over the reward.
NF2: Cross shaped non-functional	Two straight 20cm sections of aluminium, crossed to form a cross shape. Presented a more distinctive difference in shape between the functional and non-functional rake set than the non-functional rake previously presented during pilot testing (NF1, see Chapter 3).

4.2.3 Experimental procedure for those subjects who reached criterion during pilot testing (Chapter 3)

The three subjects that reached criterion during pilot testing (Win Bo, Lucia, and Truman, see Table 4.3), described in Chapter 3, were initially presented with the rake set used during pilot testing again (F1 and NF1, see section 3.2.2) for a maximum of three blocks (30 trials), assessing if they had maintained performance levels since their initial exposure. Criterion level across all experiments was set as 8/10 correct selections of the functional rake over two consecutive blocks. The experimental procedure during all experiments was as described in section 4.2.2, with the behavioural variables coded listed in Table 4.4. If subjects failed the repeat of pilot testing, they were given the same sequence of experiments as other subjects (see section 4.2). If they reached criterion, demonstrating levels of performance consistent with previous testing, they then moved to Experiment 2 (see section 4.4), and finally to Experiment 4 (see section 4.6), if criterion was reached during Experiment 2. These subjects did not participate in Experiment 3, as this experiment would have been a second repeat of pilot testing. In contrast to the main group of subjects, who were given a maximum of three blocks during Experiments 3 and 4, these previously successful subjects were given a maximum of three blocks during Experiments 2 and 4. This presentation of experiments aimed to assess if these subjects could retain performance levels when the rake end-shape was changed, with the non-functional rake shape changing in Experiment 2, and the functional rake in Experiment 4. The results from these subjects are presented in section 4.8.

Table 4.3: Subjects' species, age, sex, housing information who passed the original presentation in pilot testing (Chapter 3), details of which experiments they participated in and whether they reached criterion level performance or not.

Subject	Enclosure	Species	Age	Sex	Group	Repeat of pilot testing	Exp. 2	Exp. 4
Win Bo	6	*H. leuconedys*	12	M	Male/female pair	X[1]		
Truman	9	*H. pileatus*	12	M	Solitary	✓	✓	X
Lucia	15	*N. leucogenys*	5	F	Family	✓	X	

[1] This subject was given the same series of experiments as the main group of subjects following his failure during the repeat of pilot testing.

✓ = Reached criterion level

X = Did not reach criterion level

4.2.4 Behavioural variables

As subjects were free to move around their enclosure during testing, a 1m² area surrounding the testing apparatus was designated as the 'target area'. Subject's time spent within the target area attending to the apparatus was coded to measure the time spent engaged with the task (see Table 4.4). Time to first productive rake contact was taken from the time subjects entered the target area and attended to the apparatus, until the first rake contact that moved the rake towards the subject. This was taken as a measure of cognitive processing time, assessing how long subjects took to decide which rake to select. The time spent in reward-directed and non-reward directed rake contacts were coded, with reward directed contact being defined as any contact that pulled the rake closer to the subject. Non-directed contact was classified as contact which lasted 0.5 seconds or longer, without moving the rake towards the subject. Behavioural coding was completed using 'Solomon Coder', a video coding program freely available online (Péter, 2018). For analyses, age was categorised as adult (8+ years), or sub-adult (4-8 years). During pilot testing (see Chapter 3), age was categorised as in Palombit (1994, adult 8+ years, sub-adult 6-8 years, adolescent 3-5 years), however, adolescent and sub-adult were now pooled together as there was only one adolescent subject.

Table 4.4: Definitions of the behaviours coded from trial videos for all experiments.

Behaviour	Description
Time in target area attending to apparatus	Time spent within the designated target area visually attending to, or in physical contact with any part of the apparatus through hands, feet or mouth.
Time to first productive rake contact	Time to first contact that moved the rake towards the subject in a reward directed manner. Taken from the time subjects entered the target area and

	attended to the apparatus, until first reward directed contact.
Reward directed contact	Duration of contact with either rake that moved the rake closer to the subject.
Non-reward directed contact	Duration of contact with either rake which did not move the rake closer to the subject. Contact was classified as non-reward directed if it occurred for 0.5 seconds or longer but did not move the rake closer to the subject.

4.3 Experiment 1

4.3.1 Experimental procedure and apparatus

The first experiment presented subjects with the functional rake used during pilot testing (F1), and the cross shaped non-functional (NF2), with handles differing in colour (see Figure 4.5). The clear colour and shape changes presented here aimed to reduce the cognitive demands of the task by making the differences more obvious. The functional rake end was a rectangular shape, usable to slide the food reward into reach, whereas the cross shaped non-functional slid by the reward. The rake handles were coloured either red, or yellow and green striped, with the colours used for each rake held constant for each individual. Previous research has found gibbons can learn to associate colour cues (green, yellow and red) with hidden reward types and so it was assumed subjects would reliably differentiate between colours presented here (see D'Agostino & Cunningham, 2015). The colours used were varied randomly between subjects, but always between subjects tested in the same enclosure. For example, if two subjects were tested within an enclosure, one would be given a red handled functional rake, and green/yellow non-functional, whereas the second subject would receive a green/yellow functional and red non-functional. This aimed to minimise any effects from subjects witnessing other individuals in their enclosures completing testing.

Testing procedure was as described above (section 4.2.2) and behavioural responses were coded as shown in Table 4.4. During this experiment, subjects could be using differences in rake colour, shape, or functionality to guide their selections.

General methods were as described in section 4.2.2 above. A maximum of 12 blocks per individual (consisting of ten trials each) were presented, with subjects excluded from further experiments if criterion level was not reached. As two or more individuals were tested in enclosures 1, 6, 10, 11 (separated for testing) and 5 (tested in family group), these subjects may have witnessed other individuals in their enclosures participating in testing. Order effects were therefore analysed, with the order of testing not appearing to consistently affect performance levels (see section 4.3.2.1 for analysis of order effects).

Figure 4.5: Siamang (*S. syndactylus*) selecting the functional rake (F1), shown with a red coloured handle. The cross shaped non-functional rake (NF2), with a green and yellow striped handle, remains in the starting position.

4.3.2 Results

4.3.2.1 General performance overview

All 18 subjects reached criterion level performance (8/10 correct selections of the functional rake across two consecutive blocks) within the maximum of 12 blocks (see Table 4.5). The addition of colour and shape differences between the rakes therefore appeared to aid subjects in reliably selecting the functional rake. There were individual differences in performance across subjects, with three individuals reaching criterion within the minimum of two blocks and the remaining subjects requiring between three and eleven blocks to reach criterion level. Win Bo, who reached criterion during pilot testing (Chapter 3, section 3.3), failed to display task sufficient performance during a repeat presentation of this task and so was given the same series of experiments as the main group of subjects (see sections 4.2.3 and 4.8). Despite this additional experience of the raking in task, Win Bo did not perform better than the other subjects tested, taking the greatest number of trials to reach criterion level performance. It is possible that this subject's increased experience with the functional (F1) and non-functional (NF1) rakes presented during pilot testing interfered with learning the task using a visually different rake set; however, as Win Bo was the only gibbon to complete testing in this order, this is a speculative observation. It is also of note that this subject had ongoing health problems, which may have affected his performance.

Performance across all trials completed by each subject was evaluated using binomials tests. All subjects other than Khin Maung Win and U Myint Swe performed significantly above chance levels when all trials were considered, with these two subjects performing poorly in earlier blocks before reaching criterion level in the last two blocks (see Table 4.5). These two subjects also displayed side biases during testing, selecting the rake positioned on the same side of the testing platform across all trials during some blocks, suggesting they took longer to learn which features of the rakes to attend to in order to reliably select the functional rake.

Table 4.5: Subjects' genus, age group and sex. Number of correct selections of the functional rake in each block completed by subjects (1 block = 10 trials). Blocks where subjects displayed a side bias that potentially hindered performance or reached criterion level performance are highlighted. Significance on binomial tests for number of correct selections of the functional rake across the total number of trials completed by each subject, and during the blocks where criterion level performance was achieved are also shown.

Subject	**Genus**	**Age group**	**Sex**	**Total number of correct trials per block**											**Total number of correct trials**	**Performance on criterion blocks**
				1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**		
Arthur	*Hoolock*	Adult	M	8	8										16/20*	16/20*
Betty	*Hoolock*	Adult	F	8	9										17/20**	17/20**
Chan Thar	*Hoolock*	Adult	F	5	8	8									21/30*	16/20*
Khin Maung Win	*Hoolock*	Sub-adult	M	4	4	4	5	5	7	7	8	9			53/90	17/20**
Phy Gyi	*Hoolock*	Adult	F	5	7	9	8								29/40**	17/20**
U Myint Swe	*Hoolock*	Sub-adult	M	5	5	3	6	5	5	9	9				47/80	18/20**

Win Bo	*Hoolock*	Adult	M	5	7	7	7	4	6	7	7	7	10	8	75/110**	18/20**
Chloe	*Hylobates*	Adult	F	7	7	10	6	9	8						47/60**	17/20**
Domino	*Hylobates*	Adult	M	8	7	8	9								32/40**	17/20**
Iszie	*Hylobates*	Sub-adult	F	6	6	7	6	9	10						44/60**	19/20**
Ivan	*Hylobates*	Adult	M	5	7	7	6	10	6	9	8				58/80**	17/20**
Khusus	*Hylobates*	Adult	F	6	6	8	7	8	10						45/60**	18/20**
Reg	*Hylobates*	Adult	M	7	6	6	7	7	8	10					51/70**	18/20**
Tuk	*Hylobates*	Adult	F	8	7	8	8								31/40**	16/20*
Violet	*Hylobates*	Sub-adult	F	9	6	8	9								32/40**	17/20**
Asteriks	*Nomascus*	Adult	F	9	9										18/20**	18/20**
Ricky	*Nomascus*	Adult	F	5	9	10									24/30**	19/20**
Marlow	*Symphalangus*	Adult	F	6	6	6	6	7	6	7	9	8			61/90**	17/20**

= Side bias

= Criterion level performance

* Significant at $p < 0.05$ level

** Significant at $p < 0.01$ level

As in Chapter 3, Generalized Estimating Equations (GEE), an extension of generalized linear models, were used to build a model assessing the effect of age and sex on the number of correct selections of the functional rake when the number of blocks completed was controlled for (see Table 4.6). The model was specified as Poisson loglinear with log link function, with number of correct selections of the functional rake as the dependent variable. Age group and sex were set as factors. For parameter estimates, sub-adult was set as the reference category when evaluating age group differences, and female the reference category when evaluating sex differences. Total number of blocks completed was set as a covariate, assessing whether those subjects who required a greater number of blocks to reach criterion level performance displayed fewer selections of the functional rake per block. The working correlation matrix was specified as 'independent' meaning trials were assumed to be independent of each other. Age was categorised as adult (8+ years), or sub-adult (4-8 years). Genus was excluded from the GEE model due to there being only one *Symphalangus*, and two *Nomascus* subjects.

Table 4.6: GEE parameter estimates for factors used in model to compare frequency of functional rake selections between age groups and sexes, with number of blocks completed as a covariate.

Factor	**Parameter**	**β**	**SE**	**Lower CI**	**Higher CI**	**Wald Chi (χ^2)**	**df**	**Sig.**
	Number of blocks completed	-0.02	0.01	-0.03	-0.01	22.39	1	**<0.01**
Age group	Adult	0.10	0.04	0.02	0.18	6.56	1	**0.01**
	Sub-adult	0^a						
Sex	Male	-0.03	0.03	-0.08	0.02	1.06	1	0.30
	Female	0^a						

Significant results are in bold

[a] Set to zero because this parameter is redundant

Age was a significant predictor of performance when sex and number of blocks were held constant ($p < 0.01$), with adults selecting the functional rake more frequently than sub-adults (see Table 4.6). Sex did not significantly predict performance levels, although females performed marginally better than males (see Table 4.5). The number of blocks subjects completed before reaching criterion level significantly predicted performance levels ($p < 0.01$). Number of blocks completed was a significant predictor of performance in the expected direction, with those subjects that needed more blocks to reach criterion selecting the non-functional rake more frequently on average per block of trials (see Figure 4.6). *Hoolock* subjects took an average of 5.57 blocks to reach criterion level, with *Hylobates* subjects taking an average of 7.5 blocks. The two *Nomascus* subjects took an average of 2.5 blocks, with the sole *Symphalangus* subject requiring 9 blocks before reaching criterion level (see Table 4.5).

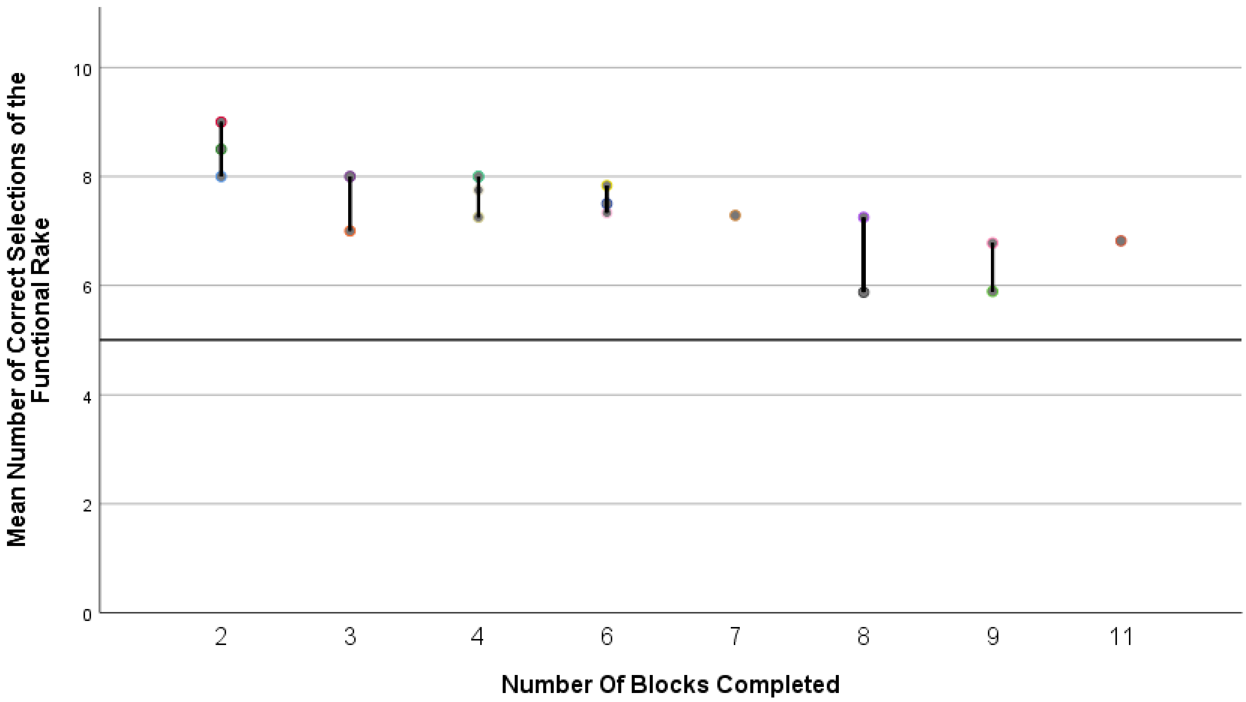


Figure 4.6: Mean number of correct selections of the functional rake (per block) for each subject who completed 2 – 11 blocks of trials, with chance level performance marked. Each coloured marker represents one subject.

4.3.2.2 Evaluation of productive behaviours

4.3.2.2.1 Time spent in target area

GEEs were used to build two models assessing the effect of age group and sex on time spent in the target area attending to the apparatus and time to first productive contact with the rakes when number of blocks completed was controlled for (see Table 4.7). Both models were specified as linear with identity link function, with either time spent attending to the apparatus or time to first productive contact as the dependent variable. Age group and sex were set as factors, with total number of blocks completed set as a covariate. For parameter estimates, sub-adult was set as the reference category when evaluating age differences, and female the reference category when evaluating sex differences. The working correlation matrix was specified as 'independent' meaning trials were assumed to be independent of each other.

Table 4.7: GEE parameter estimates for factors used in models to compare the time spent in the target area attending to the apparatus and time to first productive rake contact between age groups and sexes, with number of blocks completed as a covariate.

Model	Factor	Parameter	β	SE	Lower CI	Higher CI	Wald Chi (χ^2)	df	Sig.
Time spent in target area attending to apparatus		Number of blocks completed	-0.39	0.08	-0.55	-0.23	22.44	1	**<0.01**
	Age group	Adult	0.75	0.38	0.01	1.50	3.96	1	**<0.05**
		Sub-adult	0[a]						
	Sex	Male	0.05	0.45	-0.83	0.93	0.01	1	0.91
		Female	0[a]						
Time to first productive rake contact		Number of blocks completed	-0.29	0.07	-0.44	-0.15	15.62	1	**<0.01**
	Age group	Adult	0.63	0.38	-0.13	1.38	2.67	1	0.10
		Sub-adult	0[a]						
	Sex	Male	-0.16	0.51	-1.12	0.85	0.10	1	0.76
		Female	0[a]						

Significant results are in bold

[a] Set to zero because this parameter is redundant

Age significantly predicted time spent in the target area attending to the apparatus, ($p < 0.05$), but not time to first productive rake contact (see Table 4.7). The sub-adults (who performed poorly compared to adults), spent less time attending to the apparatus and, in general, were faster to first productive contact (mean = 2.38 seconds) than adults (mean = 3 seconds), suggesting this group spent less time considering the task before making a selection resulting in errors. Sex did not predict either time spent attending to the apparatus or time to first productive contact (see Table 4.7). Females (mean = 3.46 seconds) spent slightly less time attending to the apparatus than males (mean = 3.51 seconds), and were marginally slower to first productive contact (females mean = 2.78 seconds, males mean = 2.62 seconds).

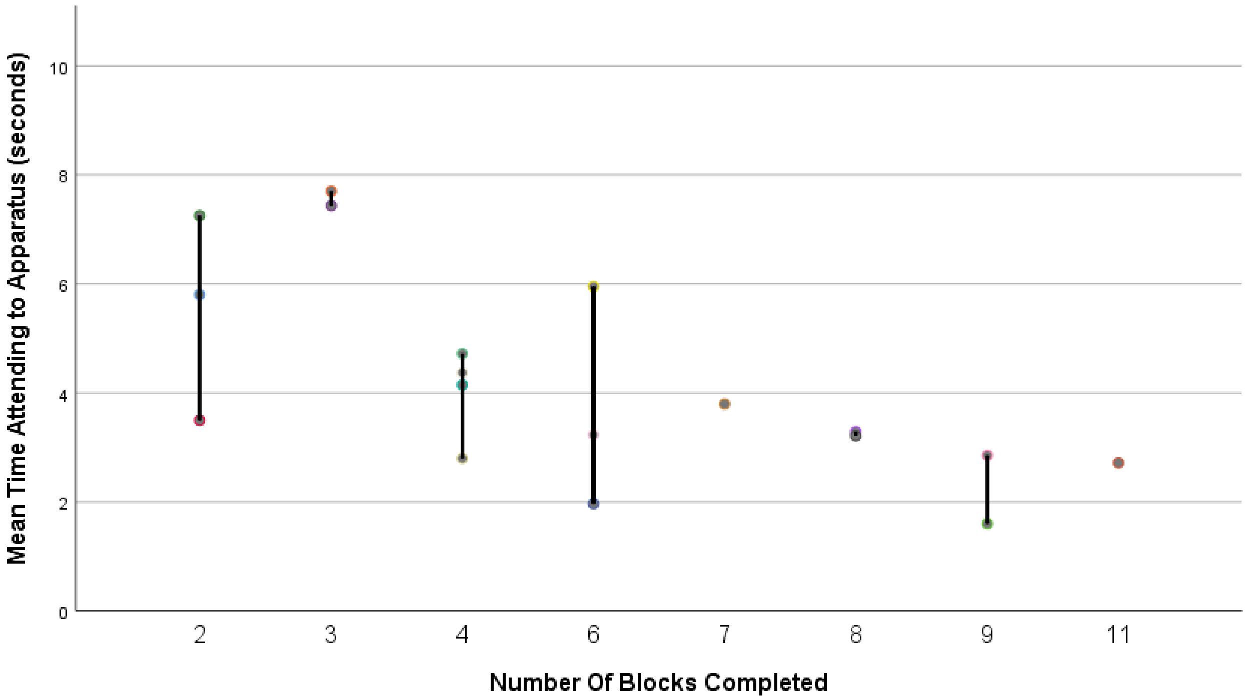


Figure 4.7: Mean time spent in the target area attending to the apparatus (per block) for each subject who completed 2 – 11 blocks of trials. Each coloured marker represents one subject.

Number of blocks completed significantly predicted time attending to the apparatus ($p < 0.01$), and time to first productive contact per block ($p < 0.01$)

(see Table 4.7). The subjects who reached criterion level performance in the fewest number of blocks spent more time attending to the task, and took longer to first contact with the rakes per block (see Figure 4.7 and 4.8), suggesting increased time spent engaged, either visually or physically, with the task aided performance.

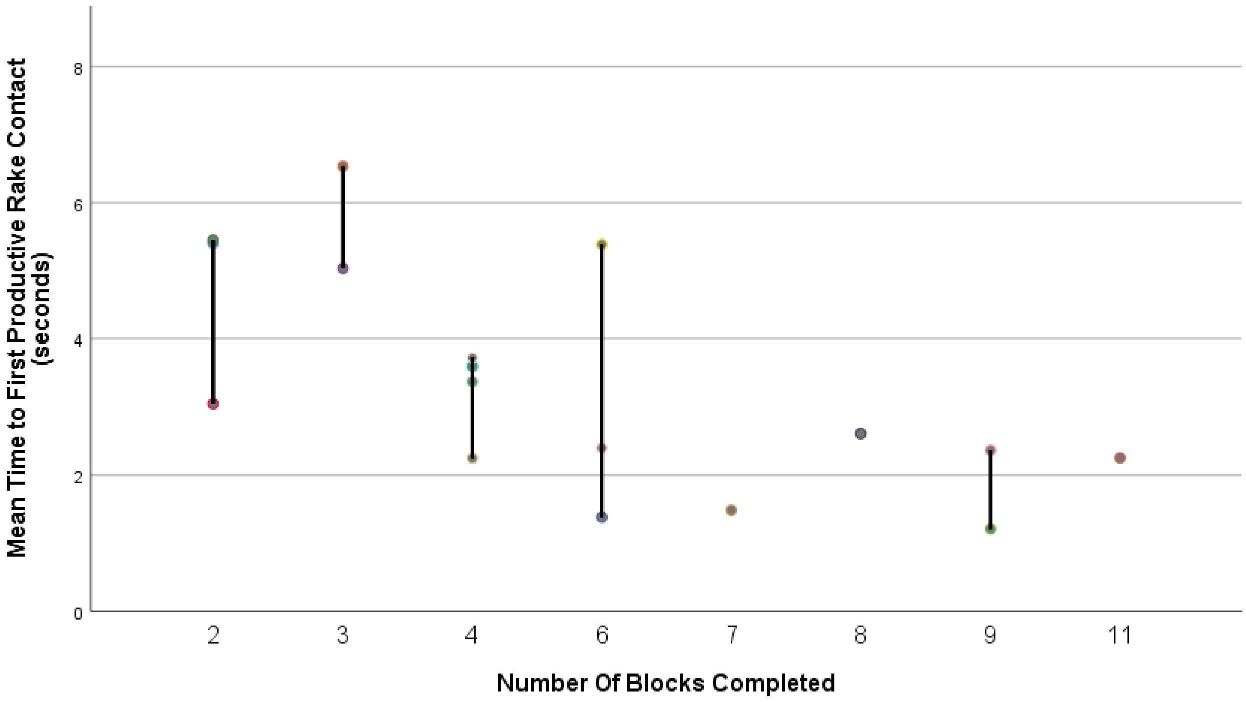


Figure 4.8: Mean time to first productive contact with the rakes (per block) for each subject who completed 2 – 11 blocks of trials. Each coloured marker represents one subject.

Although genus was not included in the model, *Nomascus* subjects spent most time attending to the apparatus, and were slowest to first productive contact, followed by *Hylobates* subjects (see Table 4.8). *Nomascus* subjects reached criterion in the fewest number of blocks, again suggesting increased time spent engaged with the task may have improved performance. *Hoolock* and the single *Symphalangus* subject spent less time attending to the task and were faster to first contact than the other genera.

Table 4.8: Mean time (seconds) spent in the target area attending to the apparatus and to first productive rake contact across all blocks completed for each genus, with standard deviations in parentheses.

Genus	Mean time attending to apparatus	Mean time to first productive contact
Hoolock (n = 7)	3.34 (2.64)	2.74 (2.42)
Nomascus (n = 2)	5.88 (7.27)	4.23 (2.09)
Hylobates (n = 8)	3.84 (2.58)	2.86 (2.44)
Symphalangus (n = 1)	2.85 (1.71)	2.36 (1.62)

4.3.2.2.2 Rake contacts

Two models were built using GEEs to assess the effect of age and sex on time spent in contact with the rakes when number of blocks completed was controlled for (see Table 4.9). The model specifications were as described above for assessment of time spent in the target area, however, time spent in reward directed contact, and time spent in non-reward directed contact were now set as the dependent variables. The model assessing time spent in non-reward directed contact produced no significant effects. As this contact type was minimal, with non-reward directed contact occurring on only 23/980 trials, this variable was excluded from all further analyses.

Table 4.9: GEE parameter estimates for factors used in model to compare time spent in reward directed rake contact between age groups and sexes, with number of blocks completed as a covariate.

Factor	Parameter	β	SE	Lower CI	Higher CI	Wald Chi (χ^2)	df	Sig.
	Number of blocks completed	-0.05	0.02	-0.09	-0.01	6.18	1	**<0.05**

Age group	Adult	-0.07	0.12	-0.31	0.17	0.32	1	0.57
	Sub-adult	0[a]						
Sex	Male	-0.03	0.12	-0.27	0.21	0.05	1	0.83
	Female	0[a]						

Significant results are in bold

[a] Set to zero because this parameter is redundant

For reward directed contact, neither age nor sex were significant predictors (Table 4.9). Despite spending less time attending to the apparatus, sub-adults spent marginally longer in reward directed contact (mean = 0.78 seconds) than adults (mean = 0.71 seconds). However, this did not result in improved performance, as adults selected the functional rake more frequently than sub-adults. Females spent slightly longer (mean = 0.76 seconds) in reward directed contact than males (mean = 0.73 seconds). Number of blocks completed significantly predicted time spent in reward directed rake contact (Table 4.9) ($p < 0.05$), with time spent per block generally decreasing as number of blocks increased (see Figure 4.9). As performance levels did not increase as the number of blocks completed increased, this suggests subjects who required more trials before reaching criterion level were less engaged with the task, spending less time attending to the apparatus and manipulating the rakes.

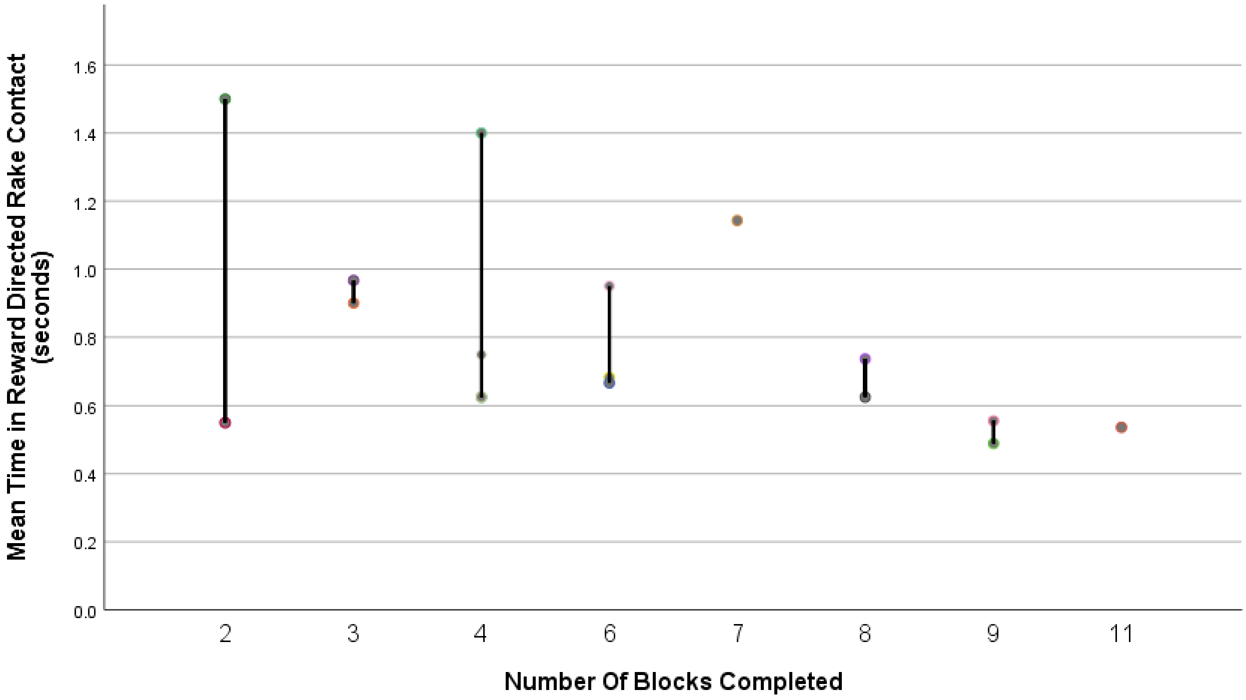


Figure 4.9: Mean time spent in reward directed rake contact (per block) for each subject who completed 2 – 11 blocks of trials. Each coloured marker represents one subject.

Nomascus subjects, who reached criterion in the fewest number of blocks, and *Hylobates* subjects, spent more time in reward directed manipulations than the other genera. These two genera also spent more time attending to the apparatus and took longest to first productive contact. The *Symphalangus* subject spent the least time engaged with the task, being fastest to first productive contact and spending the least time in reward directed rake contact.

4.3.2.3 Order effects

Order effects were assessed using Mann-Whitney U tests for all enclosures where two or more individuals participated in testing. Time to first productive contact with the rakes was compared over the first ten trials. There was a significant difference between Chloe and Ivan ($U = 23.5$, $z = -2.01$, $p < 0.05$) in enclosure 1, with Ivan (the second individual tested) being faster to first productive rake contact. Testing order did not improve Ivan’s performance however, as he selected the functional rake less frequently than Chloe (Ivan 5/10 trials, Chloe 7/10 trials). There were no significant differences between the three individuals tested in enclosure 5 (Domino and Tuk, $U = 49$, $z = -0.08$, $p = 0.97$; Tuk and Iszie, $U = 50$, $z = 0.0$, $p = 1.00$; Domino and Iszie, $U = 49$, $z = -0.08$, $p = 0.97$) or between Chan Thar and Win Bo in enclosure 6 ($U = 25$, $z = -1.89$, $p = 0.06$).

In enclosure 10, the first individual tested (U Myint Swe) took an average of 1 second to first productive contact and selected the functional rake on 5/10 trials. The second individual, Betty, selected the functional rake more frequently (8/10 trials), but was slower to first contact (mean = 5.3 seconds). Khin Maung Win, the third tested subject, was faster than Betty to first productive contact (mean = 1 second), but selected the functional rake less frequently (4/10 trials). There were significant differences between Betty and U Myint Swe ($U = 0.0$, $z = -3.78$, $p < 0.01$), and between Betty and Khin Maung Win ($U = 1$, $z = -3.71$, $p < 0.01$), but not between U Myint Swe and Khin Maung Win ($U = 27.5$, $z = -1.71$, $p = 0.09$). Being the second, or third, individual tested therefore did not appear to improve overall performance levels. Finally, in enclosure 11, there was a significant difference between Phy Gyi and Arthur ($U = 11.5$, $z = -2.91$, $p < 0.01$). Arthur, the second individual to be tested, selected the functional rake more frequently (8/10 trials) than Phy Gyi (5/10 trials), but took longer to first contact (mean = 6.8 seconds, Phy Gyi mean = 2.3 seconds), with testing order potentially affecting Arthurs performance. Although learning may have been positively affected for Arthur, both subjects reached criterion level, performing at a significant level to progress to the next experiment.

4.3.3 Summary

All subjects successfully selected the functional rake during this experiment, suggesting the increased difference in appearance between the functional and non-functional rakes aided subjects in learning to differentiate between them. The number of blocks required to reach criterion level performance varied across subjects. Those subjects who required more blocks in order to reach criterion level spent less time attending to the task and were faster to first productive rake contact, suggesting these individuals were less engaged with the task and may not have considered the consequence of their actions fully before making a choice. The increased time spent engaged with the task by subjects who reached criterion in fewer blocks suggests these subjects were considering the task before making selections, with more time available for attending to the relevant features of the different rakes.

As was found during pilot testing (Chapter 3, section 3.3), adults selected the functional rake more frequently than sub-adults, supporting previous research suggesting an increase in problem solving abilities and innovative behaviours in adult primates (Kendal, Coe & Laland, 2005; Griffin & Guez, 2014; Reader & Laland, 2001). However, the finding that adults spent significantly longer in the target area attending to the task contrasts with Kendal et al.'s (2005) finding that attentiveness levels to an extractive foraging task decreased as age increased in *Callitrichid* monkeys. Furthermore, Kendal et al. (2005) found adults had shorter latencies to first successful manipulation with the foraging task, in contrast to the current findings where adults were generally slower to first productive rake contact. It is possible differences in task types led to the differences displayed between age groups. The foraging task of Kendal et al. (2005) presents a more natural paradigm, one which adult subjects, given their life histories, would have more experience with, potentially contributing to the faster successful manipulations and decreased interest in the task found in older individuals. In contrast, the current task required object manipulation in order to retrieve an out of reach food reward, a behaviour for which there are no reports found in wild gibbons (Shumaker, Walkup & Beck, 2011). The less natural behaviour required in this task may have resulted in the delayed responses of adults in the current research, with their consideration of the task before making selections leading to improved performance levels.

Genera differences also suggested an increased time spent considering the task before making selections led to improved performance. In contrast to pilot testing, where *Nomascus* subjects displayed the lowest frequency of functional rake selections, *Nomascus* subjects now reached criterion level in the fewest number of blocks, with these subjects spending the most time attending to the task and being slowest to first productive contact with the rakes. In contrast, the single *Symphalangus* subject, who took nine blocks to reach criterion, spent the least time attending to the task and was fastest to first productive contact. There were no significant differences between the sexes; however, the general trends in performance differences were similar to those found in pilot testing. Females selected the functional rake slightly more often than males, were slower to first productive rake contact and spent longer in reward directed manipulations.

4.4 Experiment 2

4.4.1 Experimental procedure and apparatus

Experiment 2 removed the colour differences between the rakes present in Experiment 1, encouraging subjects to attend to the difference in rake end shape and functionality. As all subjects reached criterion level performance in Experiment 1, the same subjects participated in this experiment. The same rake set was presented as in Experiment 1 (F1 and NF2, see section 4.3.1) with both rakes now coloured black (see Figure 4.10). Experimental procedure was as described in section 4.2.2, with the behavioural variables coded as shown in Table 4.4. A maximum of 12 blocks were presented to each subject, with individuals excluded from further experiments if criterion level performance was not reached within these 120 trials. As subjects in enclosures 1, 6, 10, 11 (separated for testing) and 5 (tested in family group) may have witnessed other individuals in their enclosures being tested order effects were analysed (see section 4.4.2.1). Testing order did not appear to influence performance as the second, or third, tested individuals did not perform better than the subjects tested first.

Figure 4.10: Javan gibbon (*H. moloch*) selecting the functional (F1) rake, with the cross shaped non-functional (NF2) rake remaining in the starting position. Both rakes have black handles.

4.4.2 Results

4.4.2.1 General performance overview

In general, subjects reached criterion level performance (8/10 correct selections of the functional rake across two consecutive blocks) in fewer trials during Experiment 2 than Experiment 1. Of the 18 subjects tested, five reached criterion level within the first two blocks (see Table 4.10), suggesting they were transferring knowledge learned during Experiment 1 and attending to aspects of the rakes other than colour differences. Ten subjects required between three and six blocks in order to reach criterion level, with these subjects potentially attending to the more salient colour differences between the rakes during

Experiment 1, being required to learn to attend to different perceptual features during this experiment before displaying criterion level performance. The remaining three failed to reach criterion level with the maximum of 12 blocks, suggesting these subjects were attending to colour differences during Experiment 1 and found it challenging to inhibit prior learning, failing to learn which aspects of the rakes to attend to in this new task.

Performance across all trials completed by each subject was evaluated using binomials tests. Domino and Ivan did not select the functional rake significantly above chance levels across all trials, nor did they reach criterion level performance, and so were excluded from further testing. Despite performing significantly above chance levels when all trials were considered, Khusus also failed to reach criterion level and was excluded from further testing. Khin Maung Win did not select the functional rake significantly above chance levels across all trials but did reach criterion level performance and so progressed to Experiment 3 (see Table 4.10).

The different trial numbers completed by subjects during Experiment 1 did not appear to have a consistent effect on performance during Experiment 2. Of the six subjects who required more than 60 trials to reach criterion level in Experiment 1, two now reached criterion level within 30 trials and one within 40 trials. Two of the remaining three reached criterion in 60 trials, with the last failing to reach criterion level within 120 trials. The increased trial numbers therefore did not appear to improve learning within these subjects. Likewise, from the three best performing subjects in Experiment 1 who reached criterion within the minimum of 20 trials, one again reached criterion within 20 trials during Experiment 2, with the remaining two requiring 50 trials to reach criterion.

Table 4.10: Subjects' genus, age group and sex. Number of correct selections of the functional rake in each block completed by subjects (1 block = 10 trials). Blocks where subjects displayed a side bias that potentially hindered performance or reached criterion level performance are highlighted. Significance on binomial tests for number of correct selections of the functional rake across the total number of trials completed by each subject, and during the blocks where criterion level performance was achieved are also shown.

Subject	**Genus**	**Age group**	**Sex**	**Total number of correct trials per block**												**Total number of correct trials**	**Performance on criterion blocks**
				1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**		
Arthur	*Hoolock*	Adult	M	6	4	7	8	8								33/50*	16/20*
Betty	*Hoolock*	Adult	F	9	8											17/20**	17/20**
Chan Thar	*Hoolock*	Adult	F	8	10											18/20**	18/20**
Khin Maung Win	*Hoolock*	Sub-adult	M	4	6	5	6	8	8							37/60	16/20*
Phy Gyi	*Hoolock*	Adult	F	6	8	8										22/30*	16/20*
U Myint Swe	*Hoolock*	Sub-adult	M	7	8	10										25/30**	18/20**

Win Bo	*Hoolock*	Adult	M	6	10	9										25/30**	19/20**
Chloe	*Hylobates*	Adult	F	8	9											17/20**	17/20**
Domino	*Hylobates*	Adult	M	6	6	8	3	2	6	5	5	4	6	6	5	62/120	n/a
Iszie	*Hylobates*	Sub-adult	F	6	8	9										23/30**	17/20**
Ivan	*Hylobates*	Adult	M	8	6	5	6	2	7	5	5	5	6	4	7	66/120	n/a
Khusus	*Hylobates*	Adult	F	6	5	4	7	8	7	6	7	6	7	7	8	78/120**	n/a
Reg	*Hylobates*	Adult	M	7	4	6	7	8	8							40/60*	16/20*
Tuk	*Hylobates*	Adult	F	9	8											17/20**	17/20**
Violet	*Hylobates*	Sub-adult	F	7	6	5	6	10	10							44/60**	20/20**
Asteriks	*Nomascus*	Adult	F	7	7	7	9	9								39/50**	18/20**
Ricky	*Nomascus*	Adult	F	8	8											16/20*	16/20*
Marlow	*Symphalangus*	Adult	F	7	7	8	10									32/40**	18/20**

= Side bias

= Criterion level performance

* Significant at $p < 0.05$ level

** Significant at $p < 0.01$ level

As in Experiment 1, a model was built using Generalized Estimating Equations (GEE) to assess the effect of age, sex and success (reached criterion level performance or not) on the number of correct selections of the functional rake when the number of blocks completed was held constant (see Table 4.11). Model specifications were as described for assessment of performance levels in Experiment 1 (see section 4.3.2.1), with the following exceptions. Success was excluded as a factor in GEE models during Experiment 1 as all subjects reached criterion in that experiment. Given that some subjects failed to reach criterion level during Experiment 2, success was included here in order to compare performance of those who failed with those who reached criterion. For parameter estimates, ‘unsuccessful’ was the reference category for assessment of performance levels between subjects who reached criterion or not.

Table 4.11: GEE parameter estimates for factors used in model to compare frequency of functional rake selections between age groups, sexes, and subjects who were successful or unsuccessful, with number of blocks completed included as a covariate.

Factor	**Parameter**	**β**	**SE**	**Lower CI**	**Higher CI**	**Wald Chi (χ^2)**	**df**	**Sig.**
	Number of blocks completed	-0.04	0.01	-0.07	-0.02	11.70	1	**<0.01**
Age group	Adult	0.01	0.04	-0.07	0.09	0.03	1	0.87
	Sub-adult	0[a]						
Sex	Male	-0.11	0.04	-0.18	-0.04	10.23	1	**<0.01**
	Female	0[a]						
Success	Successful	-0.08	0.09	-0.26	0.10	0.74	1	0.39
	Unsuccessful	0[a]						

Significant results are in bold

[a] Set to zero because this parameter is redundant

Similar patterns were seen between age groups and sexes as in Experiment 1, with adults and females again displaying a trend for selecting the functional rake more frequently than sub-adults and males. Age did not significantly predict performance levels (see Table 4.11); however, sub-adults tended to select the functional rake less frequently, on average, than adults (see Table 4.10). Sex did predict performance levels ($p < 0.01$), with females selecting the functional rake more frequently than males (see Table 4.11). Whether subjects were successful overall (reached criterion level performance) did not significantly predict how often subjects selected the functional rake per block (Table 4.11). The difference between successful and unsuccessful subjects was small, with both groups selecting the functional rake an average of 7/10 trials per block. The three unsuccessful subjects did select the functional rake on 8/10 trials on some blocks, suggesting these subjects may have been learning which aspects of the rakes to attend to (see Table 4.10). However, they failed to display this performance across two consecutive blocks, as required by the set criterion level, and so were excluded from further testing.

The number of blocks subjects completed significantly predicted frequency of functional rake selections ($p < 0.01$) (see Table 4.11). As expected, those subjects who reached criterion level faster, completing fewer blocks, selected the functional rake, on average, more frequently per block (see Figure 4.11). All *Hoolock* subjects reached criterion level, taking an average of 3.4 blocks to do so (see Table 4.10). These subjects' performance appeared to be less affected by the removal of the colour differences than other genera. The *Hylobates* subjects that reached criterion took an average of 3.5 blocks; however, three *Hylobates* subjects failed to reach criterion, suggesting their selections during Experiment 1 may have been guided by the more salient colour differences between the rakes. The two *Nomascus* subjects were the only group who took more blocks to reach criterion level in Experiment 2 (mean = 3.5 blocks) than Experiment 1 (mean = 2.5 blocks), suggesting these subjects were also attending to colour differences in Experiment 1 and re-learning which aspects of the rakes to attend to. The *Symphalangus* subject reached criterion level in fewer blocks than Experiment 1; however, this subject performed poorest, requiring four blocks to reach criterion.

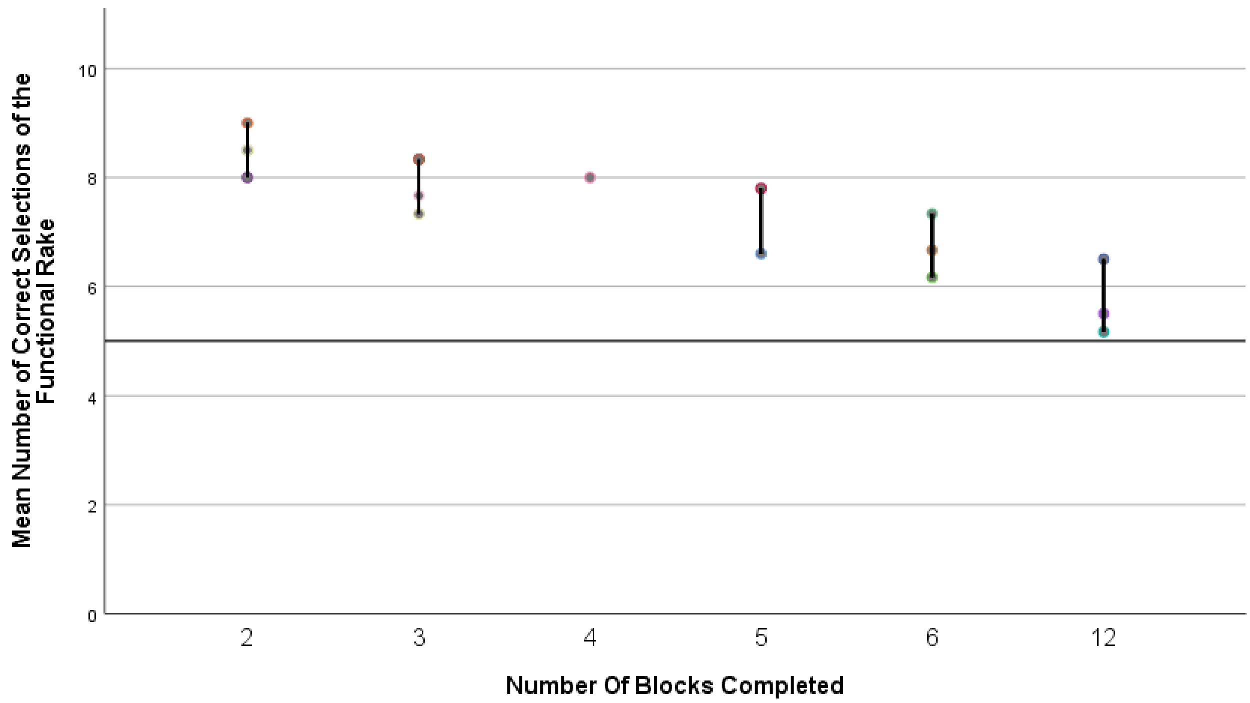


Figure 4.11: Mean number of correct selections of the functional rake (per block) for each subject who completed 2 – 12 blocks of trials, with chance level performance marked. Each coloured marker represents one subject.

4.4.2.2 Evaluation of productive behaviours

4.4.2.2.1 Time spent in the target area

GEEs were used to build two models assessing the effect of age group, sex and success (reached criterion level or not) on time spent in the target area attending to the apparatus and time to first productive rake contact when the number of blocks completed was controlled for (see Table 4.12). Both models were specified as for assessment of time spent in the target area during Experiment 1 (see section 4.3.2.2.1) with the following exception. Success was now included as a factor, with ‘unsuccessful’ set as the reference category for parameter estimates.

Table 4.12: GEE parameter estimates for factors used in models to compare time spent in target area attending to the apparatus and time to first productive rake contact between age groups, sexes, and subjects who were successful or unsuccessful, with number of blocks completed as a covariate.

Model	Factor	Parameter	β	SE	Lower CI	Higher CI	Wald Chi (χ^2)	df	Sig.
Time spent in target area attending to apparatus		Number of blocks completed	-0.08	0.23	-0.53	0.37	0.12	1	0.73
	Age group	Adult	1.80	0.50	0.82	2.78	13.01	1	**<0.01**
		Sub-adult	0[a]						
	Sex	Male	-0.09	0.52	-1.11	0.93	0.03	1	0.86
		Female	0[a]						
	Success	Successful	1.18	1.60	-1.96	4.31	0.54	1	0.46
		Unsuccessful	0[a]						
Time to first productive rake contact		Number of blocks completed	-0.27	0.25	-0.77	0.23	1.11	1	0.29
	Age group	Adult	1.22	0.69	-0.14	2.57	3.09	1	0.08
		Sub-adult	0[a]						
	Sex	Male	-0.26	0.55	-1.34	0.87	0.21	1	0.64

		Female	0[a]						
	Success	Successful	-0.89	2.05	-4.91	3.13	0.19	1	0.67
		Unsuccessful	0[a]						

Significant results are in bold

[a] Set to zero because this parameter is redundant

Age significantly predicted time spent attending to the apparatus only ($p < 0.01$) (see Table 4.12). As in Experiment 1, adults spent longer attending to the task, and generally were slower to first productive contact (mean = 3.05 seconds) than sub-adults (mean = 1.84 seconds), with more time engaged with the task, and slower more considered behaviour potentially leading to the better performance levels in this group. Sex was not a significant predictor in either model (see Table 4.12). In contrast to Experiment 1, females generally spent longer attending to the apparatus than males (females mean = 2.82 seconds, males mean = 2.72 seconds). Females were again slower to first contact (females mean = 2.57 seconds, males mean = 2.32 seconds), as was found in the first experiment. Success was also not a significant predictor in either model (see Table 4.12). Subjects who reached criterion level tended to spend longer attending to the task in each block completed (successful mean = 3.36 seconds, unsuccessful mean = 2.18 seconds). However, successful subjects were faster to first contact (mean = 2 seconds) than unsuccessful subjects (mean = 2.89 seconds).

Number of blocks completed produced a more variable pattern. The best performing subjects (reaching criterion within the minimum of two blocks) generally spent more time attending to the apparatus and took longer to first productive contact per block. However, the number of blocks subjects required to reach criterion did not significantly predict either time spent attending to the apparatus or time to first productive contact (see Table 4.12). *Hoolock* subjects again spent less time attending to the task and were faster to first productive rake contact than either *Symphalangus* or *Nomascus* subjects (see Table 4.13). In contrast to Experiment 1, where *Nomascus* subjects reached criterion level in the fewest number of blocks, *Hoolock* subjects reached criterion level faster in Experiment 2. This suggests more focussed, efficient processing in these gibbons that could have been a direct result of previous experience with the task in the previous experiment. As in Experiment 1, *Nomascus* subjects spent longest attending to the task and were slowest to first productive rake contact. *Nomascus* subjects and successful *Hylobates* subjects took the same number of blocks to reach criterion. However, *Hylobates* spent the least time engaged with the task and were faster to first productive contact, suggesting *Nomascus* subjects required longer processing time before making the correct selection.

Table 4.13: Mean time (seconds) spent in the target area attending to the apparatus and to first productive rake contact across all blocks for each genus, with standard deviations in parentheses.

Genus	Mean time attending to apparatus	Mean time to first productive contact
Hoolock (n =7)	3.81 (2.75)	3.24 (2.65)
Nomascus (n = 2)	4.18 (1.79)	3.67 (1.70)
Hylobates (n = 8)	3.12 (2.04)	2.32 (1.82)
Symphalangus (n = 1)	4.07 (1.72)	3.36 (1.68)

4.4.2.2.2 Rake contacts

A model was built using GEEs to assess the effect of age, sex and success (reached criterion level performance or not) on time spent in reward directed contact with the rakes when number of blocks completed was controlled for (see table 4.14). The model specifications were as described for assessment of time spent in rake contact during Experiment 1 (see section 4.3.2.2.2), with the exception that success was now included as a factor. For parameter estimates, ‘unsuccessful’ was again set as the reference category. As in Experiment 1, non-reward directed rake contact was minimal and excluded from analyses.

Table 4.14: GEE parameter estimates for factors used in model to compare time spent in reward directed rake contact between age groups, sexes, and subjects who were successful or unsuccessful, with number of blocks completed as a covariate.

Factor	Parameter	β	SE	Lower CI	Higher CI	Wald Chi (χ^2)	df	Sig.
	Number of blocks completed	0.02	0.04	-0.06	0.09	0.15	1	0.70
Age group	Adult	0.06	0.11	-0.16	0.29	0.32	1	0.57
	Sub-adult	0[a]						
Sex	Male	-0.06	0.07	-0.20	0.09	0.61	1	0.43
	Female	0[a]						
Success	Successful	0.27	0.37	-0.46	1.01	0.53	1	0.47
	Unsuccessful	0[a]						

[a] Set to zero because this parameter is redundant

Neither age nor sex were significant predictors of time spent in reward directed contact with the rakes (see Table 4.14). As in Experiment 1, females (mean = 0.64 seconds), who selected the functional rake more frequently, spent slightly longer in reward directed contact than males (mean = 0.58 seconds). In contrast to Experiment 1, the adult group (mean = 0.64 seconds) spent longer in reward directed contact than sub-adults (mean = 0.58 seconds). Success was also not a significant predictor (see Table 4.14); however, subjects who reached criterion level generally spent longer in reward directed rake manipulations (mean = 0.75 seconds, unsuccessful subjects mean = 0.47 seconds). Time spent in rake contact per block varied across subjects, with number of blocks completed not significantly predicting time in reward directed contact (see Table 4.14).

Despite spending the longest time attending to the apparatus, and being slowest to first productive contact, *Nomascus* subjects spent the least time in reward directed manipulations, suggesting these subjects made focussed manipulations once they had considered which rake to select. *Hoolock* subjects, who reached criterion in the fewest number of blocks, spent the second least time in rake manipulations, again suggesting more focussed behaviours by these subjects. *Hylobates* subjects, who spent the least time attending to the task but reached criterion level in the same number of blocks as *Nomascus* subjects, spent the second longest in reward directed manipulations, suggesting differences in approach to the task between *Nomascus* and *Hylobates* subjects. In contrast to Experiment 1, where the *Symphalangus* subject spent the least time attending to the task and in reward directed manipulations, this subject spent the most time in rake manipulations during Experiment 2. This subject also took the highest number of blocks to reach criterion (when compared to the mean number of blocks needed by other genera), suggesting the removal of colour differences may have resulted in this subject having to relearn which aspects of the rakes to attend to, resulting in slower processing.

4.4.2.3 Order Effects

Order effects were assessed using Mann-Whitney U tests for all enclosures where two or more individuals participated in testing. Time to first productive contact with the rakes was compared over the first ten trials. There was a significant difference between Chloe and Ivan in enclosure 1, ($U = 17$, $z = -2.5$, $p = 0.01$). Ivan was second to be tested, and was faster to make first productive contact with the rakes (mean = 4.46 seconds) than Chloe (mean = 6.35 seconds). However, both subjects selected the functional rake on the same number of trials (8/10), suggesting testing order did not improve overall performance. There were no significant differences between the three subjects tested in enclosure 5 (Domino and Tuk $U = 33$, $z = -1.29$, $p = 0.22$; Tuk and Iszie $U = 37$, $z = -0.99$, $p = 0.35$; Domino and Iszie $U = 44$, $z = -0.46$, $p = 0.68$). In enclosure 6, there was a significant difference between Chan Thar and Win Bo, the second individual to be tested ($U = 0.0$, $z = -3.79$, $p < 0.01$). Win Bo was faster to first productive contact with the rakes (mean = 1.34 seconds, Chan Thar mean = 7.45 seconds), however, he selected the functional rake less frequently than Chan Thar (Win Bo 6/8 trials, Chan Thar 8/10 trials), again suggesting testing order did not improve performance.

In enclosure 10, the first subject tested (U Myint Swe) took an average of 0.93 seconds to first productive contact and selected the functional rake on 7/10 trials. Khin Maung Win, the second tested, was both slower to first productive contact (mean = 1.82 seconds), and selected the functional rake less frequently (4/10 trials). Betty, who was tested last, selected the functional rake most frequently (9/10 trials), but was slowest to first productive contact (mean = 6.04 seconds). The differences between Khin Maung Win and Betty ($U = 4$, $z = -3.48$, $p < 0.01$), and U Myint Swe and Betty ($U = 0.0$, $z = -3.79$, $p < 0.01$) were significant, whereas the difference between U Myint Swe and Khin Maung Win was not ($U = 31$, $z = -1.44$, $p = 0.17$). Being the second, or third, individual tested therefore did not appear to improve overall performance. Finally, in enclosure 11, where both subjects selected the functional rake on 6/10 trials, there was a significant difference between Phy Gyi and Arthur ($U = 6.5$, $z = -3.29$, $p < 0.01$). Arthur, the second individual tested, was slower to first productive contact (mean = 6.47 seconds, Phy Gyi mean = 2.55 seconds), again suggesting testing order did not provide an advantage to subjects' overall performance.

4.4.2.4 Comparison of performance in Experiments 1 and 2

Wilcoxon signed ranks tests were used to assess the difference in performance levels between the last ten trials of Experiment 1, and the first ten trials of Experiment 2, investigating subjects' initial performance in response to the removal of the colour differences between the rakes. There was a significant difference in frequency of functional rake selections, with subjects selecting the functional rake more frequently during the last ten trials of Experiment 1 than the first trials of Experiment 2 ($Z = -3.25$, N for Test = 14, $p < 0.01$). Subjects therefore appeared to be attending to the more salient colour differences between the rakes during Experiment 1, with the removal of this perceptual difference causing performance levels to drop as subjects re-learned which aspects of the rakes to attend to.

This suggestion is supported by the finding that subjects spent longer attending to the apparatus (Experiment 2 mean = 4.34 seconds, Experiment 1 mean = 4.26 seconds), and were slower to first productive contact (Experiment 1 mean = 3.25 seconds, Experiment 2 mean = 3.49 seconds) during the first ten trials of Experiment 2, with this increased time considering the task before making selections suggesting they were re-learning which aspects of the rakes were important. That subjects reached criterion level in fewer blocks overall during Experiment 2 than Experiment 1, suggests they were able to learn which aspects to attend to faster, potentially transferring knowledge of rake end shape or functionality from Experiment 1. This is supported by the finding that subjects spent significantly less time in reward directed contact during the first ten trials of Experiment 2 ($Z = -3.29$, N for Test = 147, $p < 0.01$), suggesting more focussed manipulations after considering the task.

4.4.3 Summary

The current findings suggest gibbons learning was not tied to specific perceptual features, with 15 subjects continuing to select the functional rake above chance level once colour differences were removed. This finding is in line with Fujita et al. (2003), who found three of four tested capuchins (*C. apella*) continued to select a functional rake significantly above chance levels when the rake colours were changed. Subjects did appear to transfer some knowledge of either rake-end shape or differences in functionality from Experiment 1, requiring fewer blocks before reaching criterion level during Experiment 2. However, the initial drop in performance levels, and that subjects spent longer considering the task before making selections, suggested subjects could have been attending to the colour differences during Experiment 1, with the removal of this more salient difference requiring re-learning of which properties of the rakes to attend to. Three of the 18 subjects tested failed to reach criterion, suggesting these individuals were using colour differences between the rakes during Experiment 1 for success. Performance differences between successful and unsuccessful subjects were small, and it is possible these unsuccessful subjects may have learned to attend to either rake-end shape or functionality, given sufficient experience.

The general trend in age group differences were similar to Experiment 1, with adults tending to select the functional rake more frequently, spending longer engaged with the task and being slower to first productive rake contact. General trends in sex differences were also similar to Experiment 1, with females performing better than males, spending longer engaged with the task and being slower to first contact. There were no significant differences between successful and unsuccessful subjects; however, successful subjects did generally spend longer attending to the task and in reward directed rake manipulations. Taken together these findings suggest slower, more considered behaviour led to improved performance.

Hoolock subjects' performance appeared to be the least affected by the removal of the colour differences between the rakes, with all subjects reaching criterion level, and doing so in fewer blocks than Experiment 1. As in Experiment 1, these subjects spent generally spent less time engaged with the task before making selections, suggesting more focussed and efficient processing. In contrast, three *Hylobates* subjects failed to reach criterion, and *Nomascus* subjects, despite all reaching criterion, were the only group to take more blocks to reach criterion level during Experiment 2 than Experiment 1. The better performance shown by *Hoolock* subjects here supports findings of Cunningham (2006), who found *Hoolock* subjects consistently performed better than other genera (*Nomascus*, *Hylobates* and *Symphalangus*) during object manipulation and tool use tasks.

Hylobates subjects spent the least, and *Nomascus* subjects the most, time attending to the task and before making first productive rake contact. The successful subjects from these groups took the same number of blocks to reach criterion, suggesting differences in approach to the task between these genera, with *Nomascus* subjects spending longer considering the task before making selections. In contrast to Experiment 1, where the single *Symphalangus* subject spent the least time engaged with the task and was fastest to first productive contact, she now displayed slower behaviour than both *Hoolock* and *Hylobates* subjects. This subject reached criterion level in fewer blocks in Experiment 2 suggesting the increased attentiveness to the apparatus, potentially as a result of the removal of the more obvious perceptual colour differences, improved her performance.

4.5 Experiment 3

4.5.1 Experimental procedure and apparatus

Presenting visually different, but functionally similar rake sets allows assessment of whether subjects learned to attend to specific perceptual features such as rake-end shape rather than functionality to make selections. If subjects were able to generalise knowledge and had attended to functionality across previous experiments, performance levels ought to remain high when rake end-shape was changed. However, if subjects' selections were guided by rake-end shape (rather than functionality), it is likely performance levels would be poorer when rake shapes were novel as subjects would have to re-learn the correct response. Experiment 3 therefore presented those subjects that had reached criterion during Experiment 2 with the functional rake used during Experiments 1 and 2, and a rectangular non-functional rake (F1 and NF1, see Figure 4.12), used during pilot testing (see Chapter 3, section 3.2.2). All but one of the subjects tested here failed to reach criterion level performance on first presentation of this rake (see Chapter 3, section 3.3). This allowed assessment of whether subjects were now able to reliably select the functional rake after learning which aspects of the rakes to attend to during Experiments 1 and 2.

The maximum number of blocks given to subjects was limited to three, in order to assess initial performance following the change in rake end-shape and ability to generalise across functional shapes rather than ability to learn this new configuration over repeated presentations. Any subject not reaching criterion level performance during these three blocks was excluded from further experiments. Experimental procedure was as described in section 4.2.2, with the behavioural variables coded as shown in Table 4.4. Subjects in enclosures 6, 10, 11 (separated for testing) and 5 (tested in family group) may have witnessed other individuals in their enclosures participating in testing. Order effects were therefore analysed, with the order of testing not found to consistently improve performance levels (see section 4.5.2.1).

Figure 4.12: Pileated gibbon (*H. pileatus*) selecting the rectangular non-functional rake (NF1), with the functional rake (F1) remaining in the starting position.

4.5.2 Results

4.5.2.1 General performance overview

Ten subjects reached criterion level performance (8/10 correct selections of the functional rake across two consecutive blocks) within the maximum of three blocks in Experiment 3 (see Table 4.15). Five subjects failed to reach criterion. These subjects potentially attended to the shape of the non-functional rake during Experiment 2, rather than understanding the properties that made it non-functional, and failed to learn to differentiate between the rakes presented here, which were more similar in appearance. Performance across all trials completed by each subject was evaluated using binomials tests. Arthur, Phy Gyi, and Tuk did not select the functional rake significantly above chance levels, nor did they reach criterion level performance, and so were excluded from further testing. Chloe and Reg did perform significantly above chance levels across all trials, but failed to reach criterion level and so did not progress to Experiment 4. Overall performance was improved from pilot testing (see Chapter 3, section 3.3), where only one of these subjects reached criterion, compared to ten subjects here, suggesting some subjects had learned which aspects of the rakes to attend to during Experiments 1 and 2, potentially having some understanding of functionally relevant properties.

The different trial numbers given during Experiment 2 did not appear to have a consistent effect on performance during Experiment 3. Of the five subjects who reached criterion in the minimum of 20 trials during Experiment 2, two again reached criterion level within 20 trials. Of the remaining three, two failed to reach criterion level and one did so in 30 trials. Likewise, for those three subjects who completed the most trials (60 trials) during Experiment 2; two now reached criterion within 20 trials, with the third failing to reach criterion level. Completing more trials during Experiment 2 could have resulted in increased depth of learning; however, this did not appear to be the case for all subjects as effects were inconsistent.

Table 4.15: Subjects' genus, age group and sex. Number of correct selections of the functional rake across the total number of trials completed by each subject, with significance on binomial test shown (trials were presented in blocks of 10), and whether subjects reached criterion level performance.

Subject	Genus	Age Group	Sex	Total number of correct trials	Reached criterion level performance
Arthur	*Hoolock*	Adult	M	20/30	X
Betty	*Hoolock*	Adult	F	16/20*	✓
Chan Thar	*Hoolock*	Adult	F	16/20*	✓
Khin Maung Win	*Hoolock*	Sub-adult	M	16/20*	✓
Phy Gyi	*Hoolock*	Adult	F	18/30	X
U Myint Swe	*Hoolock*	Sub-adult	M	21/30*	✓
Win Bo	*Hoolock*	Adult	M	18/20**	✓
Chloe	*Hylobates*	Adult	F	22/30*	X
Iszie	*Hylobates*	Sub-adult	F	24/30**	✓
Reg	*Hylobates*	Adult	M	23/30**	X
Tuk	*Hylobates*	Adult	F	20/30	X
Violet	*Hylobates*	Sub-adult	F	19/20**	✓
Asteriks	*Nomascus*	Adult	F	22/30*	✓
Ricky	*Nomascus*	Adult	F	26/30**	✓
Marlow	*Symphalangus*	Adult	F	18/20**	✓

* Significant at $p < 0.05$ level

** Significant at $p < 0.01$ level

A model was built used Generalized Estimating Equations (GEE) to assess the effect of age, sex and success (reached criterion level performance or not) on the number of correct selections of the functional rake when the number of blocks completed was controlled for (see Table 4.16). Model specifications were as described for assessment of performance levels in Experiment 2 (see section 4.4.2.1).

Table 4.16: GEE parameter estimates for factors used in model to compare frequency of functional rake selections between age groups, sexes, and subjects who were successful or unsuccessful, with number of blocks completed as a covariate.

Factor	**Parameter**	**β**	**SE**	**Lower CI**	**Higher CI**	**Wald Chi (χ^2)**	**df**	**Sig.**
Age group	Adult	0.02	0.06	-0.09	0.12	0.07	1	0.79
	Sub-adult	0[a]						
Sex	Male	-0.01	0.05	-0.10	0.08	0.03	1	0.86
	Female	0[a]						
Success	Successful	0.13	0.07	-0.00	0.26	3.59	1	0.06
	Unsuccessful	0[a]						

[a] Set to zero because this parameter is redundant

None of the factors were found to significantly predict performance levels (see Table 4.16). As in previous experiments, adults displayed a trend for performing better than sub-adults (see Table 4.15). Males and females selected the functional rake at the same frequency. As expected, but in contrast to Experiment 2, subjects who reached criterion showed a slight increase in functional rake selections per block (mean = 8/10 trials), than subjects who did not reach criterion level (mean = 7/10 trials). The single *Symphalangus* subject reached criterion level within 2 blocks, faster than she had reached criterion during Experiments 1 or 2, suggesting experience gained during these experiments led her to attend to relevant features of the rake ends faster (see Table 4.15). The remaining three genera displayed the same order of performance as Experiment 2, with the successful *Hoolock* subjects reaching criterion level within an average of 2.2 blocks, the successful *Hylobates* subjects within 2.5 blocks, and the *Nomascus* subjects within 3 blocks. All successful subjects reached criterion level in fewer blocks than required during previous experiments, suggesting they were transferring learning across the experiments, potentially attending to the functional features of the rakes faster during Experiment 3.

4.5.2.2 Evaluation of productive behaviours

4.5.2.2.1 Time spent in the target area

GEEs were used to build two models assessing the effect of age group, sex and success (reached criterion level or not) on time spent in the target area attending to the apparatus and time to first productive rake contact when number of blocks completed was controlled for (see Table 4.17). Both model specifications were as described for assessment of time spent in the target area during Experiment 2 (see section 4.4.2.2.1).

Table 4.17: GEE parameter estimates for factors used in models to compare the time spent in the target area attending to the apparatus and time to first productive rake contact between age groups, sexes, and subjects who were successful or unsuccessful, with number of blocks completed as a covariate.

Model	Factor	Parameter	β	SE	Lower CI	Higher CI	Wald Chi (χ^2)	df	Sig.
Time spent in target area attending to apparatus	Age group	Adult	1.60	0.83	-0.04	3.23	3.66	1	0.06
		Sub-adult	0[a]						
	Sex	Male	-0.44	0.80	-2.02	1.14	0.30	1	0.59
		Female	0[a]						
	Success	Successful	1.22	0.98	-0.70	3.13	1.56	1	0.21
		Unsuccessful	0[a]						
Time to first productive rake contact	Age group	Adult	1.54	0.80	-0.04	3.11	3.65	1	0.06
		Sub-adult	0[a]						
	Sex	Male	-0.40	0.80	-1.96	1.16	0.26	1	0.61
		Female	0[a]						
	Success	Successful	1.29	0.89	-0.46	3.04	2.09	1	0.15
		Unsuccessful	0[a]						

[a] Set to zero because this parameter is redundant

None of the factors significantly predicted either time attending to the apparatus or time to first contact, however, the general patterns of performance across the two models were similar (see Table 4.17). As in previous experiments, adults tended to spend longer attending to the task (mean = 3.77 seconds) and were slower to first productive contact (mean = 3.08 seconds) than sub-adults (time attending mean = 2.18, time to first contact mean = 1.55 seconds). Females generally spent more time attending to the task than males (females mean = 3.19 seconds, males mean = 2.75 seconds), and were slower to first productive contact (females mean = 2.52 seconds, males mean = 2.11 seconds), a trend seen across previous experiments.

Subjects who reached criterion tended to spend longer attending to the apparatus (mean = 3.58 seconds, unsuccessful mean = 2.36 seconds) and were slower to first productive contact (mean = 2.96 seconds, unsuccessful mean = 1.67 seconds). This trend contrasts with Experiment 2, where there was little difference in average frequency of functional rake selections between successful and unsuccessful subjects, but successful subjects were slightly faster to first productive contact. The change in rake-end shape during Experiment 3 may have resulted in increased difficulty discriminating between the rakes (when compared to the more obvious difference in Experiment 2) with those subjects who spent longer attending to the rakes before making selections achieving slightly higher frequencies of functional rake selections.

Genera differences were the same as in Experiment 2 (see Table 4.18). *Hylobates* subjects again spent the least time attending to the task and were fastest to first productive contact, with this more impulsive behaviour perhaps leading to the lower levels of functional rake selections shown by these subjects. *Hoolock* subjects spent less time attending to the task and were faster to first productive rake contact than *Symphalangus* or *Nomascus* subjects. In contrast to *Hylobates* subjects, this faster behaviour may represent more efficient processing, as *Hoolock* subjects reached criterion level in fewer blocks than *Nomascus* or *Hylobates* subjects. The *Symphalangus* subject was slower to first productive rake contact and spent longer attending to the apparatus than either *Nomascus* or *Hoolock* subjects. However, in contrast to Experiment 2, this resulted in an improved performance by this subject, who now reached criterion within the minimum of two blocks. *Nomascus* subjects, who took longest to reach criterion level performance, spent more time attending to the task and to first contact, suggesting these subjects were re-learning which aspects of the rake to attend to before reaching task sufficient performance.

Table 4.18: Mean time (seconds) spent in the target area attending to the apparatus and to first productive rake contact across all blocks for each genus, with standard deviations in parentheses.

Genus	Mean time attending to apparatus	Mean time to first productive contact
Hoolock (n = 7)	3.48 (3.04)	2.88 (2.84)
Nomascus (n = 2)	4.81 (1.75)	4.31 (1.65)
Hylobates (n = 5)	3.14 (1.88)	2.33 (1.75)
Symphalangus (n = 1)	3.74 (1.28)	2.94 (1.26)

4.5.2.2.2 Rake contacts

A model was built using GEEs to assess the effect of age, sex and success (reached criterion level performance or not) on time spent in reward directed contact with the rakes when number of blocks completed was controlled for (see Table 4.19). The model specifications were as described for assessment of time spent in rake contact during Experiment 2 (see section 4.5.2.2.2). As in previous experiments, non-reward directed rake contact was minimal and excluded from analyses.

Table 4.19: GEE parameter estimates for factors used in model to compare time spent in reward directed rake contact between age groups, sexes, and subjects who were successful or unsuccessful, with number of blocks completed as a covariate.

Factor	Parameter	β	SE	Lower CI	Higher CI	Wald Chi (χ^2)	df	Sig.
Age group	Adult	0.02	0.10	-0.18	0.21	0.02	1	0.89
	Sub-adult	0[a]						
Sex	Male	-0.13	0.10	-0.32	0.07	1.55	1	0.21
	Female	0[a]						
Success	Successful	-0.10	0.13	-0.36	0.16	0.57	1	0.45
	Unsuccessful	0[a]						

[a] Set to zero because this parameter is redundant

None of the factors were significant predictors of time spent in reward directed contact with the rake (see Table 4.19). Adults (mean = 0.69 seconds), spent marginally longer in reward directed rake contact than sub-adults (mean = 0.68 seconds); however this difference was minimal. In line with previous experiments, females (mean = 0.75 seconds) spent longer in reward directed rake contact than males (mean = 0.63 seconds), though not significantly so (see Table 4.19). As a group, and in contrast to Experiment 2, those subjects who reached criterion level generally spent less time in reward directed contact (mean = 0.64 seconds) than unsuccessful subjects (mean = 0.74 seconds).

Time spent in reward directed contact showed a similar pattern across genera as in Experiment 2, with *Hoolock* and *Nomascus* subjects spending less time in reward directed contact compared to the *Hylobates* subjects or the single *Symphalangus*. *Hylobates* spent the longest in reward directed manipulations, despite spending the least time attending to the task before making selections. In contrast, *Nomascus* subjects, who took more blocks to reach criterion level than only *Hylobates* subjects, spent the least time in reward directed contact. *Nomascus* subjects also spent the longest attending to the task before making selection, suggesting these subjects were re-learning which aspects of the rakes to attend to during Experiment 3.

4.5.2.3 Order effects

Mann Whitney U tests were used to analyse order effects for all enclosures where two or more individuals participated in testing, comparing time to first productive rake contact over the first ten trials. There was no significant difference between Iszie and Tuk in enclosure 5 ($U = 44.5$, $z = -0.42$, $p = 0.68$). In enclosure 6, there was a significant difference between Win Bo and Chan Thar ($U = 7$, $z = -3.26$, $p < 0.01$). Chan Thar, the second individual to be tested, took longer to first productive contact (mean = 3.23 seconds, Win Bo mean = 1.24 seconds) and selected the functional rake less frequently (8/10 trials, Win Bo 10/10 trials). Potentially witnessing Win Bo completing trials therefore did not improve Chan Thar's performance.

In enclosure 10, the first individual tested (Betty) took longest to first productive contact (mean = 8.34 seconds) and selected the functional rake on 8/10 trials. U Myint Swe, the second to be tested, was faster to first productive contact (mean = 1.89 seconds) than Betty, but selected the functional rake less frequently (5/10 trials). Khin Maung Win was tested last, taking a similar amount of time to U Myint Swe to first productive contact (mean = 1.8 seconds) but selecting the functional rake more frequently (8/10 trials). The differences between Betty and U Myint Swe ($U = 4.5$, $z = -3.44$, $p < 0.01$), and Betty and Khin Maung Win ($U = 2$, $z = -3.63$, $p < 0.01$) were significant. There was no significant difference between the U Myint Swe and Khin Maung Win ($U = 47.5$, $z = -0.19$, $p = 0.85$). Testing order therefore did not appear to improve overall performance in this enclosure. Finally, in enclosure 11, there was a significant difference between Phy Gyi and Arthur ($U = 0.0$, $z = -3.79$, $p < 0.01$). The second individual to be tested (Arthur) selected the functional rake more frequently (8/10 trials, Phy Gyi 5/10 trials) but was slower to first make productive contact (mean = 6.11 seconds) than Phy Gyi (mean = 1.65 seconds). Although testing order may have affected Arthur's performance, this did not result in a greatly improved performance as neither individual in this enclosure reached criterion level.

4.5.2.4 Comparison of performance in Experiments 2 and 3

Wilcoxon signed ranks tests were used to assess the difference in performance levels between the last ten trials of Experiment 2 and the first ten trials of Experiment 3. This assessed whether the change in non-functional rake negatively impacted initial performance levels, suggesting subjects were re-learning which aspects of the rakes were functionally important and had been attending to shape differences in Experiment 2. Only those subjects who participated in both experiments were included in analyses.

There was a significant difference in selections of the functional rake between the last ten trials of Experiment 2 and the first ten trials of Experiment 3 ($Z = -2.73$, N for Test = 10, $p < 0.01$). Subjects selected the functional rake more frequently during the last ten trials of Experiment 2, with performance levels dropping during the initial trials of Experiment 3. Although this suggests subjects were re-learning which aspects of the rakes to attend to during Experiment 3, there was also a significant difference in time spent attending to the apparatus ($Z = -2.8$, N for Test = 144, $p < 0.01$), and time to first productive rake contact ($Z = -2.15$, N for Test = 148, $p < 0.05$). In contrast to the comparison between Experiments 1 and 2, subjects spent less time attending to the apparatus, and were faster to first productive rake contact during the first ten trials of Experiment 3 than the last trials of Experiment 2. Subjects also spent significantly less time in reward directed rake contact during the first ten trials of Experiment 3 ($Z = -2.14$, N for Test = 126, $p < 0.05$).

Given the decreased time spent attending to the apparatus, and faster time to first productive contact, subjects appeared to be less engaged with the task following the change in non-functional rake shape, something which may have led to the decreased in frequency of functional rake selections during the initial trials. The decline in interest may have been due to previous experience with this rake set during pilot testing (see Chapter 3). Given that all but one of the subjects had previously failed to reliably select the functional rake during this experimental set-up, this experience may have led to decreased interest in the task as they had learned that rewards were not regularly attainable with these apparatus (due to their lack of understanding of the task demands) . However, ten subjects did now display criterion level performance during Experiment 3, suggesting they transferred knowledge of functionally relevant features learned across Experiments 1 and 2 in order to succeed with this rake set.

4.5.3 Summary

Ten of the 15 subjects reached criterion level performance in Experiment 3, nine of which had previously failed to reliably select the functional rake during this experimental set-up (see Chapter 3, section 3.3). The findings of Experiment 3 suggest subjects had learned to attend to either the shape or functionality of the functional rake presented across Experiments 1 and 2, and transferred this knowledge to Experiment 3. The experience gained across previous experiments therefore aided subjects in succeeding with the experimental set-up during Experiment 3, where the rewards position relative to the rakes, either positioned in front of the rakes or not, could not be used to guide selections. This task presented a partial replication of the inverted rake task developed by Povinelli and Reaux (2000). Chimpanzees (*P. troglodytes*), vervet monkeys (*C. aethiops*) and tamarins (*S. oedipus*) have been found to display performance close to chance levels during assessments of these species spontaneous understanding of the task (Povinelli & Reaux, 2000; Santos et al., 2006). The current findings suggest that gibbons can also succeed at this task, given an extended learning period and experience of differentiating between rakes with more salient perceptual differences in Experiments 1 and 2. As gibbons displayed evidence of re-learning which aspects of the rakes to attend to during the initial trials of this experiment, it is likely gibbons success was not based on an understanding of the functionality of the rakes, but rather they transferred some knowledge of attending to the rake ends from Experiment 2, and were highly efficient in learning which aspects of the novel rake to attend to. Those subjects who failed to reach criterion level may have been attending to either the shape of the non-functional rake or the position of the food rewards during previous experiments, being unable to inhibit prior learning quickly enough to learn new associations, resulting in the poor performance shown here.

Although non-significant, the general trends in performance differences between age groups and sexes were similar to previous experiments. Adults tended to select the functional rake more frequently than sub-adults and displayed slower behaviour, spending longer attending to the task, in reward directed manipulations and before making first productive contact with the rakes. Females also generally displayed slower behaviour than males, as across previous experiments, with longer latencies to first contact, and longer attending to the task and in rake manipulations. The differences in performance between successful and unsuccessful subjects were also non-significant. Successful subjects again tended to spend longer attending to the apparatus and manipulating the rakes, but were now marginally slower to first productive contact, something which may have resulted in the slightly improved performance of successful subjects when compared to Experiment 2.

The change in shape of the non-functional rake did not appear to affect the *Symphalangus* subjects' performance, as she reached criterion within the minimum of two blocks, suggesting she had transferred knowledge of the functional rake from previous experiments. The pattern of performance between the remaining genera was similar to that in Experiment 2, with *Hoolock* subjects reaching criterion level in the fewest blocks, followed by *Hylobates* and *Nomascus* subjects. *Hoolock* and *Hylobates* subjects spent less time attending to the apparatus, and were faster to first productive rake contact than the worse performing *Nomascus* subjects, suggesting *Nomascus* subjects were not transferring learning to the same degree as other genera, needing to re-learn which aspects of the rakes to attend to during Experiment 3.

4.6 Experiment 4

4.6.1 Experimental procedure and apparatus

Following from Experiment 3, which presented a change in shape of the non-functional rake, Experiment 4 changed the shape of the functional rake, presenting a novel thin functional rake (F2), together with a cross shaped non-functional rake (NF2, see Figure 4.13). This assessed whether subjects learning across previous experiments could be transferred to a novel tool, or if performance had been due to an association with the functional rake-end shape. The maximum number of blocks was again limited to three per subject, assessing subjects' initial performance in response to the change in functional rake-end shape, rather than assessing their learning across repeated presentations. As such, if subjects' had been attending solely to the shape of the functional rake during previous experiments, it was expected performance levels would be lower on first exposure to this new rake. In contrast, if subjects were capable of transferring knowledge of functionality across rake sets, it was expected they would reliably select the new shaped functional rake.

The ten subjects that reached criterion on Experiment 3 participated in Experiment 4. Testing procedure was as described in section 4.2.2, with the coded behavioural variables listed in Table 4.4. Subjects in enclosures 6 and 10 were separated for testing but may have witnessed other individuals in their enclosures participating in testing. Order effects were therefore analysed, with testing order not influencing performance (see section 4.6.2.1 for analysis of order effects).

Figure 4.13: Pileated gibbon (*H. pileatus*) selecting the cross shaped non-functional rake (NF2), with the thin functional rake (F2) remaining in the starting position.

4.6.2 Results

4.6.2.1 General performance overview

Eight of the ten subjects reached criterion level performance (8/10 correct selections of the functional rake across two consecutive blocks) within the maximum of three blocks during Experiment 4 (see Table 4.20), suggesting they were capable of transferring knowledge of functionality across rake sets. Performance across all trials completed by each subject was evaluated using binomials tests, with all subjects performing significantly above chance levels. Khin Maung Win and Ricky failed to reach criterion level, despite selecting the functional rake significantly above chance levels when all trials were considered, suggesting these subjects' selections across previous experiments may have been guided by the functional rake-end shape rather than an understanding of functionality.

Table 4.20: Subjects' genus, age group and sex. Number of correct selections of the functional rake across the total number of trials completed by each subject, with significance on binomial tests shown (trials were presented in blocks of 10), and whether subjects reached criterion level performance.

Subject	**Genus**	**Age group**	**Sex**	**Total number of correct trials**	**Reached criterion level performance**
Betty	*Hoolock*	Adult	F	19/20**	✓
Chan Thar	*Hoolock*	Adult	F	26/30**	✓
Khin Maung Win	*Hoolock*	Sub-adult	M	21/30*	X
U Myint Swe	*Hoolock*	Sub-adult	M	22/30*	✓
Win Bo	*Hoolock*	Adult	M	19/20**	✓
Iszie	*Hylobates*	Sub-adult	F	19/20**	✓
Violet	*Hylobates*	Sub-adult	F	20/20**	✓
Asteriks	*Nomascus*	Adult	F	18/20**	✓
Ricky	*Nomascus*	Adult	F	23/30**	X
Marlow	*Symphalangus*	Adult	F	18/20**	✓

* Significant at $p < 0.05$ level

** Significant at $p < 0.01$ level

A model was built used Generalized Estimating Equations (GEE) to assess the effect of age, sex and success (reached criterion level performance or not) on the number of correct selections of the functional rake when the number of blocks completed was controlled for (see Table 4.21). Model specifications were as described for assessment of performance levels in Experiment 2 (see section 4.4.2.1).

Table 4.21: GEE parameter estimates for factors used in model to compare frequency of functional rake selections between age groups, sexes, and subjects who were successful or unsuccessful, with number of blocks completed as a covariate.

Factor	Parameter	β	SE	Lower CI	Higher CI	Wald Chi (χ^2)	df	Sig.
Age group	Adult	0.01	0.03	-0.05	0.07	0.11	1	0.75
	Sub-adult	0[a]						
Sex	Male	-0.07	0.04	-0.15	0.01	2.93	1	0.09
	Female	0[a]						
Success	Successful	0.09	0.03	0.03	0.15	7.87	1	**<0.01**
	Unsuccessful	0[a]						

Significant results are in bold

[a] Set to zero because this parameter is redundant

Neither age nor sex significantly predicted performance, with adults and sub-adults selecting the functional rake at the same levels (see Table 4.20). In line with Experiments 1 and 2, females tended to perform better than males, although not significantly so (see Table 4.21). Success (whether subjects reached criterion level or not) significantly predicted performance ($p < 0.01$), with successful subjects selecting the functional rake more frequently per block, as expected (see Table 4.21). Four of the five *Hoolock* subjects reached criterion level, doing so in an average of 2.5 blocks. In contrast to Experiments 2 and 3, where *Hoolock* subjects reached criterion level in fewer blocks than *Hylobates* or *Nomascus* subjects, *Hoolock* subjects now took the greatest number of blocks to reach criterion. *Hoolock* subjects were also the only genera to take more blocks to reach criterion level during Experiment 4 than Experiment 3, suggesting these subjects may have been attending to the functional rake shape during previous experiments. The two *Hylobates* subjects took an average of 2 blocks to reach criterion. The single *Nomascus* subject who reached criterion also took 2 blocks, as did the single *Symphalangus* subject (see Table 4.20).

4.6.2.2 Evaluation of productive behaviours

4.6.2.2.1 Time spent in the target area

GEEs were used to build two models assessing the effect of age group, sex and success (reached criterion level or not) on time spent in the target area attending to the apparatus and time to first productive rake contact when number of blocks completed was controlled for (see Table 4.22). Both model specifications were as described for assessment of time spent in the target area during Experiment 2 (see section 4.4.2.2.1).

Table 4.22: GEE parameter estimates for factors used in models to compare the time spent in the target area attending to the apparatus and time to first productive rake contact between age groups, sexes, and subjects who were successful or unsuccessful, with number of blocks completed as a covariate.

Model	Factor	Parameter	β	SE	Lower CI	Higher CI	Wald Chi (χ^2)	df	Sig.
Time spent in target area attending to apparatus	Age group	Adult	0.89	0.56	-0.21	1.99	2.53	1	0.11
		Sub-adult	0[a]						
	Sex	Male	-3.47	4.32	-4.32	-2.62	64.43	1	**<0.01**
		Female	0[a]						
	Success	Successful	1.11	0.38	0.36	1.85	8.55	1	**<0.01**
		Unsuccessful	0[a]						
Time to first productive rake contact	Age group	Adult	0.76	0.52	-0.27	1.79	2.11	1	0.15
		Sub-adult	0[a]						
	Sex	Male	-3.21	0.49	-4.16	-2.25	43.47	1	**<0.01**
		Female	0[a]						
	Success	Successful	0.88	0.35	0.20	1.56	6.36	1	**<0.05**
		Unsuccessful	0[a]						

Significant results are in bold

[a] Set to zero because this parameter is redundant

Age was not a significant predictor in either model (see Table 4.22). As was found across previous experiments, although non-significant, adults generally spent longer attending to the apparatus (mean = 4.12 seconds, sub-adults mean = 3.23 seconds) and took longer to first productive contact with the rakes (mean = 3.38 seconds, sub-adults mean = 2.62 seconds). Sex was a significant predictor of both time spent attending to the apparatus ($p < 0.01$) and time to first contact ($p < 0.01$) (see Table 4.22). Females spent longer attending to the apparatus and were slower to first contact, as was the general trend across Experiments 2 and 3. Success was also a significant predictor across both models, with successful subjects spending longer engaged with the task, again in line with the general trend found during Experiment 3 (see Table 4.22). Those subjects who reached criterion level spent longer attending to the task ($p < 0.01$), and were slower to first productive contact ($p < 0.05$), suggesting this increased time spent considering the task improved performance.

As in previous experiments, *Nomascus* subjects spent the longest time attending to the apparatus, but were slowest to first productive contact (see Table 4.23). *Hylobates* subjects spent the second longest attending to the task and to first contact, followed by *Hoolock* subjects, who took longer to reach criterion level performance than the other genera, and finally the *Symphalangus* subject who was the fastest.

Table 4.23: Mean time (seconds) spent in the target area attending to the apparatus and to first productive rake contact across all blocks for each genus, with standard deviations in parentheses.

Genus	Mean time attending to apparatus	Mean time to first productive contact
Hoolock (n = 5)	4.49 (3.28)	3.61 (2.84)
Nomascus (n = 2)	5.01 (2.04)	4.49 (1.84)
Hylobates (n = 2)	4.94 (2.85)	4.12 (2.84)
Symphalangus (n = 1)	3.65 (1.24)	2.84 (1.19)

4.6.2.2.2 Rake contacts

A model was built using GEEs to assess the effect of age, sex and success (reached criterion level performance or not) on time spent in reward directed contact with the rakes when number of blocks completed was controlled for (see Table 4.24). The model specifications were as described for assessment of time spent in rake contact during Experiment 2 (see section 4.5.2.2.2). Non-reward directed contact was again minimal, and so excluded from analyses.

Table 4.24: GEE parameter estimates for factors used in model to compare time spent in reward directed rake contact between age groups, sexes, and subjects who were successful or unsuccessful, with number of blocks completed as a covariate.

Factor	**Parameter**	**β**	**SE**	**Lower CI**	**Higher CI**	**Wald Chi (χ^2)**	**df**	**Sig.**
Age group	Adult	0.05	0.14	-0.22	0.33	0.16	1	0.69
	Sub-adult	0[a]						
Sex	Male	-0.17	0.85	-0.33	0.00	3.81	1	0.05
	Female	0[a]						
Success	Successful	0.21	0.05	0.11	0.31	16.51	1	**<0.01**
	Unsuccessful	0[a]						

Significant results are in bold

[a] Set to zero because this parameter is redundant

Neither age nor sex were significant predictors of time spent in reward directed contact (see Table 4.24). As in Experiments 2 and 3, adults (mean = 0.72 seconds) did spend marginally longer in effective rake contact than sub-adults (mean = 0.66 seconds), however, this did not result in an improved performance as adults and sub-adults selected the functional rake at equivalent levels. Although non-significant, females (mean = 0.77 seconds) spent longer in rake contact than males (mean = 0.61 seconds), as in all previous experiments. Success significantly predicted time spent in reward-directed contact ($p < 0.01$) (see Table 4.24). In line with the general trends in previous experiments, subjects who were successful spent longer in rake contact, suggesting increased time spent engaged with the task improved performance during this experiment.

As in Experiments 2 and 3, *Hylobates* and *Symphalangus* subjects spent longer in reward directed contact than *Nomascus* or *Hoolock* subjects, with *Hoolock* the only genus to take longer than two blocks to reach criterion level. The single *Symphalangus* subject spent the most time in reward directed contact, despite spending the least time attending to the task and to first productive contact. In contrast, *Nomascus* subjects again spent the least time in reward directed rake contact after spending longest attending to the task and to first contact, suggesting consideration of the task before making selections by these subjects.

4.6.2.3 Order effects

Mann Whitney U tests were used to analyse order effects for enclosures where two or more individuals participated in testing. Time to first productive contact with the rakes over the first ten trials was compared. There was a significant difference between Win Bo and Chan Thar in enclosure 6 ($U = 9$, $z = -3.11$, $p < 0.01$). Chan Thar, the second individual to be tested, was both slower to first productive contact (mean = 6.77 seconds, Win Bo mean = 2.42 seconds) and selected the functional rake on fewer trials (7/10, Win Bo 9/10) suggesting testing order did not provide a performance advantage. In enclosure 10, the first individual tested (U Myint Swe) took an average of 1.85 seconds to first productive contact and selected the functional rake least frequently (5/10 trials). Betty, the second individual tested, was slowest to first contact (mean = 5.29 seconds), but selected the functional rake most frequently (9/10 trials). The last individual tested, Khin Maung Win, was fastest to first contact (mean = 1.76 seconds), selecting the functional rake on 6/10 trials. There was a significant difference between U Myint Swe and Betty ($U = 9$, $z = -3.11$, $p < 0.01$), and between Betty and Khin Maung Win ($U = 8$, $z = -3.19$, $p < 0.01$). There was no significant difference between the U Myint Swe and Khin Maung Win ($U = 34$, $z = -1.22$, $p = 0.25$). Being the second, or third, individual tested therefore did not appear to improve overall performance.

4.6.2.4 Comparison of performance in Experiments 3 and 4

Wilcoxon signed ranks tests were used to assess the difference in performance levels between the last ten trials of Experiment 3 and the first ten trials of Experiment 4, investigating whether subjects' performance levels dropped following the change in functional rake shape, suggesting previous selections had been guided by rake-end shape. Only those subjects who participated in both experiments were included in analyses. Subjects selected the functional rake more frequently during the last ten trials of Experiment 3, however, the difference between Experiment 3 (mean = 8.6/10 trials) and 4 (mean = 8.2/10 trials) was not significant. There was, however, a significant difference in the time spent attending to the apparatus ($Z = -3.58$, N for Test = 99, $p < 0.01$), with subjects spending more time in Experiment 4. Subjects also took significantly longer to first productive rake contact in the first 10 trials of Experiment 4 than the last ten trials of Experiment 3 ($Z = -2.85$, N for Test = 98, $p < 0.01$) and spent significantly longer in reward directed contact ($Z = -3.3$, N for Test = 82, $p < 0.01$). Taken together these findings suggest subjects selections may have been guided by functional rake-end shape across Experiments 1-3, with the increased time spent engaged with the task and to first productive contact during the initial trials of Experiment 4 suggesting subjects were considering the apparatus before making selections. However, the finding that frequency of functional rake selections did not significantly drop suggests subjects learned quickly which aspects of the rakes to attend to.

4.6.3 Summary

Eight of the ten subjects reached criterion level in Experiment 4, suggesting, as has been found for capuchin monkeys (*C. apella*) (Fujita et al., 2003), gibbons' learning was not tied to specific perceptual features of the rakes, transferring knowledge of functionality across experiments with changing rake shapes. Age differences were non-significant; however, as in previous experiments, adults generally took longer to first productive rake contact and spent longer engaged with the task than sub-adults. Likewise, females again tended to display more considered behaviours. Unlike Experiment 3, differences between successful and unsuccessful subjects were significant, with successful subjects being slower to first productive rake contact, spending longer attending to the task and in reward directed manipulations. Slower, more considered behaviour overall again appeared to lead to improved performance, suggesting that although subjects appeared to transfer some knowledge of rake functionality from previous experiments, the change in functional rake shape required subjects to re-learn which aspects of this novel rake to attend to.

The *Symphalangus* subject again reached criterion within the minimum of two blocks, with successful *Nomascus* and *Hylobates* subjects reaching criterion level faster in Experiment 4 than Experiment 3. This suggests these subjects were transferring knowledge across experiments, rather than re-learning which aspects of the rakes to attend to. In contrast, *Hoolock* subjects took longer to reach criterion level during Experiment 4. These subjects may have been attending to the shape of the functional rake to guide selections across previous experiments, taking slightly longer to learn the correct response during Experiment 4.

4.7 Comparison of performance across Experiments 1-4

Subjects' performance during the last ten trials of each experiment were compared to the first ten trials of the following experiment. As such, only subjects who reached criterion level are considered in the last ten trials of each experiment, in order to compare these successful subjects response to the changing dimensions during the initial trials of the following experiment. There was increased variability between subjects' performance during the first ten trials of each experiment when compared to the last ten trials (see Figure 4.14), suggesting individual variation in cognitive abilities, as some individuals either transferred knowledge across the rake sets more readily, or displayed a faster processing time in learning which features of the rake to attend to. Selection of the functional rake during the first ten trials of each experiment increased as testing progressed (see Table 4.25). However, correct selections were generally lower during the first ten trials of each experiment when compared to the last ten trials of the previous experiment, suggesting the changing dimensions of the rake required subjects to re-learn the correct response with each change (see Figure 4.14). The drop in performance generally became smaller as testing progressed, suggesting subjects were transferring knowledge gained from previous experiments and learning which novel aspects of the rakes to attend to more quickly on repeated presentations.

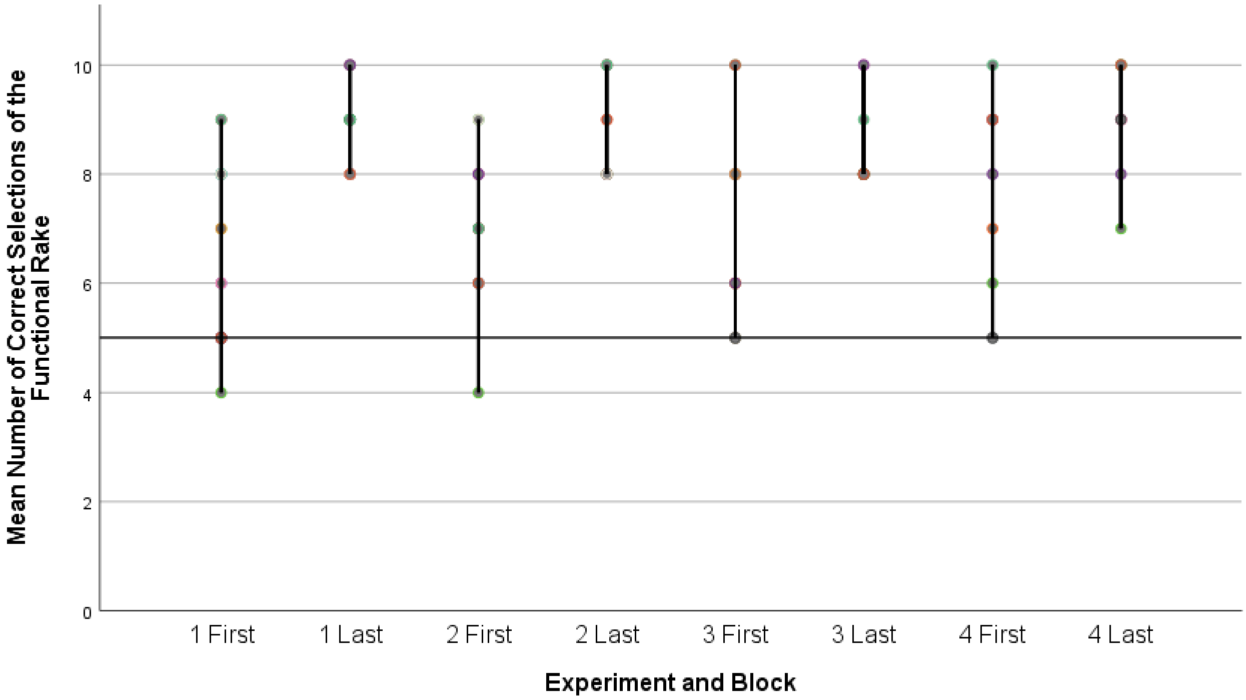


Figure 4.14: Mean number of correct selections of the functional rake across the first and last ten trials of each experiment, with chance level performance marked. The last ten trials of each experiment includes only those subjects who reached criterion level performance and progressed to the subsequent experiment. Each coloured marker represents one subject.

Time spent attending to the apparatus and time to first productive rake contact increased during the first ten trials of Experiments 2 and 4, when compared to the last ten trials of the previous experiments, again suggesting the changing dimensions of the rakes required more consideration before a selection was made (see Table 4.25). However, the time spent attending to the task, and to first productive contact decreased from the last ten trials of Experiment 2 to the first ten trials of Experiment 3, which presented the rake set used during pilot testing (see Chapter 3). The presentation of the reward situated in front of both rakes during this experiment may have posed a more difficult task than Experiments 1 and 2, where the rewards position relative to the rakes may have been used to guide selections. This increased difficulty, in combination with the previous experience of failing to learn to select the functional rake during pilot testing, frequently leading to subjects being unrewarded, may have led to decreased interest in this task during the initial trials of Experiment 3. Despite selecting the functional rake most frequently during the first ten trials of Experiment 4, subjects spent the longest time attending to the apparatus, took longest to first productive rake contact and spent longer in rake contact during these trials when compared to the first ten trials of previous experiments. The presentation of a novel functional rake in this experiment potentially imposed a higher cognitive load, forcing subjects to spend longer considering the task before making a selection. As the same functional rake was used throughout Experiments 1-3, it is possible subjects had learned to attend to the features of this rake, and were required to re-learn which aspects of the new rake to attend to.

Table 4.25: Mean time (seconds) spent attending to the apparatus, to first productive contact, and in reward directed rake contact for the first block (10 trials) and last block (10 trials) of each experiment, with standard deviations in parentheses. The last 10 trials of each experiment includes only those subjects who reached criterion level performance and progressed to the subsequent experiment.

Experiment Number	Block	Mean time attending to apparatus	Mean time to first productive contact	Mean time spent in reward directed rake contact
1	First	4.31 (2.72)	3.52 (2.57)	0.77 (0.47)
	Last	4.26 (2.68)	3.25 (2.24)	0.82 (0.46)
2	First	4.34 (2.63)	3.49 (2.48)	0.69 (0.28)
	Last	4.12 (2.49)	3.34 (2.44)	0.76 (0.35)
3	First	3.78 (2.77)	3.11 (2.63)	0.70 (0.30)
	Last	3.59 (2.38)	2.92 (2.27)	0.70 (0.33)
4	First	4.72 (3.00)	3.85 (2.70)	0.81 (0.44)

	Last	4.66 (2.88)	3.86 (2.55)	0.80 (0.33)

4.8 Results for subjects reaching criterion level performance during pilot testing

The three subjects who had previously reached criterion level performance were initially presented with a repeat of pilot testing, before completing Experiment 2 and then Experiment 4, with subjects required to reach criterion level before moving to the next experiment (see section 4.2.3). Performance across all trials completed by each subject was evaluated using binomials tests. Win Bo failed to reach criterion with the original rake set presented during pilot testing and therefore returned to Experiment 1, completing the experiments in the same sequence as other subjects who failed pilot testing (see section 4.2). Truman and Lucia both selected the functional rake significantly above chance levels across all trials and reached criterion within two blocks, progressing to Experiment 2 to assess if these subjects could transfer their learning to a novel rake set (see Table 4.26).

Lucia failed to significantly select the functional rake or reach criterion level performance during Experiment 2, and so was excluded from further testing. This subject may have been attending to the shape of the non-functional rake to make selections during the pilot testing task, rather than attending to the relevant features of both rakes, and therefore failed to transfer understanding of functionality to the new rake shape presented during Experiment 2. Truman performed significantly above chance, again reaching criterion within two blocks, and progressed to Experiment 4. During Experiment 4, Truman failed to reach criterion, selecting the functional rake around chance levels (see Table 4.26). In contrast to Lucia, Truman may have been attending to the shape of the functional rake to guide selections during the repeat of pilot testing and in Experiment 2 and therefore failed to transfer knowledge of rakes functionality to the rake set presented during Experiment 4.

Table 4.26: Number of correct selections of the functional rake per block completed by subjects who originally passed pilot testing (Chapter 3) (1 block = 10 trials), and across the total number of trials completed by each subject for each experiment (significance on binomial test). Blocks where subjects reached criterion level performance are highlighted.

Subject	Genus	Repeat of pilot testing				Experiment 2				Experiment 4			
		1	2	3	Total	1	2	3	Total	1	2	3	Total
Win Bo	*Hoolock*	6	8	6	12/30								
Truman	*Hylobates*	9	8		17/20**	8	8		16/20*	7	4	5	16/30
Lucia	*Nomascus*	8	9		17/20**	6	5	8	19/30				

= Side bias

= Criterion level performance

* Significant at $p < 0.05$ level

** Significant at $p < 0.01$ level

For both subjects who passed the repeat presentation of the pilot apparatus, the difference in time taken to first productive contact with the rakes between the last ten trials of the repeat presentation of pilot testing, and the first ten trials of Experiment 2 was not significant (Wilcoxon signed ranks test: Z = -1.01, N for Test = 20, p = 0.31). Truman displayed a greater increase than Lucia in time to first productive contact during the first trials of Experiment 2, compared to the last trials of the previous experiment, suggesting he was spending longer considering the task before making selections (see Figure 4.15). This may have led to the poorer performance displayed by Lucia, as she failed to attend to the new shape of the non-functional rake. The difference in time to first productive contact between the last ten trials of Experiment 2 and the first ten trials of Experiment 4 for Truman was also not significant (Z = 0.0, N for Test = 10, p = 1.0). Truman generally spent less time before first productive contact during the initial trials of Experiment 4, suggesting he failed to attend to the differences between the rakes when presented with a novel functional rake.

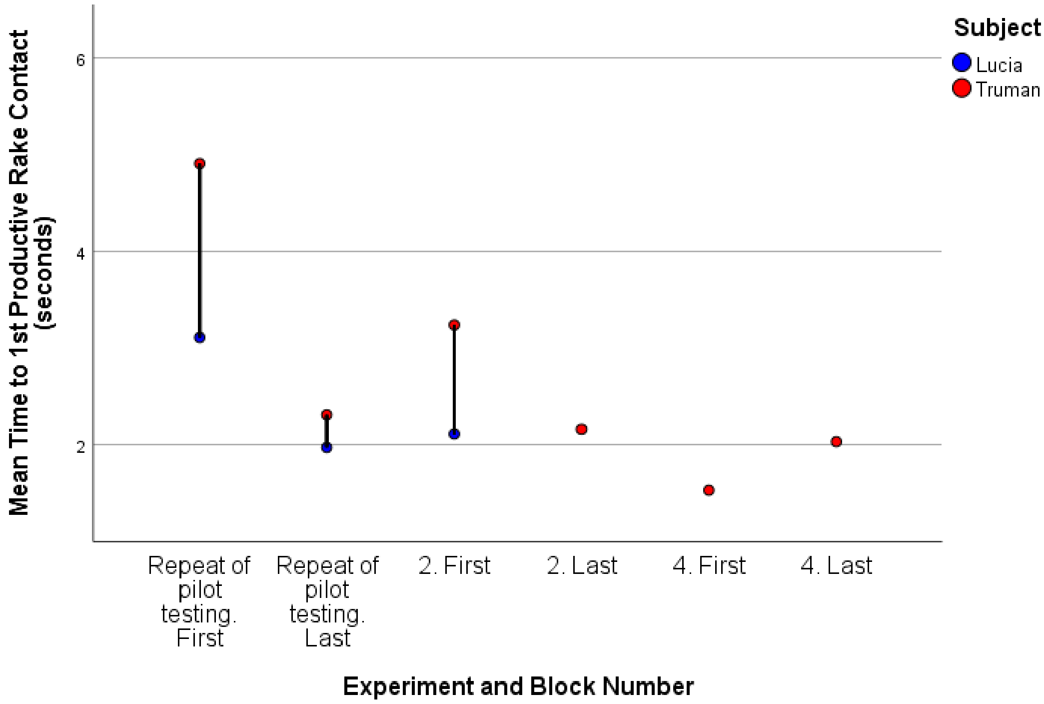


Figure 4.15: Mean time to first productive contact with the rakes per block across the first and last ten trials of the three experiments. The last ten trials of Experiment 2 include Truman only as Lucia did not progress to Experiment 4.

4.9 General discussion

Taken together, the findings suggest gibbons were capable of attending to functionally relevant features of the rakes and transferring this knowledge across different rake sets, as has been found for various species of New World monkeys (Cummins-Sebree & Fragaszy, 2005; Fujita et al., 2003; Hauser, 1997; Hauser, et al., 2002, Spaulding & Hauser, 2005). All subjects reached criterion during Experiment 1, suggesting the differences between rakes were obvious enough for subjects to differentiate between them. During this experiment subjects could have attended to differences in colour, rake-end shape, or functionality between the rakes. These findings contrasted with those of pilot testing (see Chapter 3, section 3.3), where subjects may have struggled to learn which aspects of the rakes to attend to and generally performed poorly.

The majority of subjects continued to significantly select the functional rake during Experiment 2, once the colour difference between rakes was removed. There was a drop in performance levels during the initial trials of this experiment, suggesting successful individuals may have attended to the more salient perceptual difference between the rakes (colour) during Experiment 1. Once colour differences were removed, these subjects may have been required to re-learn which aspects of the rake to attend to during Experiment 2. However, successful subjects generally reached criterion level performance in fewer blocks in Experiment 2, suggesting they had transferred some knowledge of either rake shape or functionality from Experiment 1. Given that there was some evidence of transferring learning from Experiment 1, these findings suggested that gibbons' learning was not tied to specific perceptual features of the rakes, as has been found for capuchins (*C. apella*) who continued to reliably select a functional rake once rake colours were changed (Fujita et al., 2003). Three subjects appeared to have primarily learned to attend to the colour differences between the rakes, and failed to reach criterion in Experiment 2 once this difference was removed, and only shape and functionality remained.

Experiments 3 and 4 manipulated rake shape, aiming to assess whether subjects selections were due to an association with a specific shape, or if they could transfer learning across rake sets of different shapes. The majority of subjects reached criterion level during Experiment 3, presenting the same rake set used during pilot testing, with the non-functional rake a different shape to that presented during Experiments 1 and 2. Findings again suggested subjects were transferring learning across rake sets, with experience gained across Experiments 1 and 2 aiding subjects in learning to attend to the relevant features during this task. Lucia and Truman, who reached criterion level during pilot testing, also maintained previous performance levels during this experiment, suggesting these subjects had learned to attend to either shape or functionality differences between these rakes during earlier testing and retained this knowledge.

Experiment 3 presented a partial replication of Povinelli and Reaux's (2000) inverted rake task previously presented during pilot testing, where the majority of subjects failed to reliably select the functional rake (see Chapter 3, section 3.3). Assessments of chimpanzees (*P. troglodytes*), vervet monkeys (*C.*

aethiops) and tamarins (*S. oedipus*) spontaneous understanding of this task have found these species perform around chance levels (Povinelli & Reaux, 2000; Santos et al., 2006). Six gibbons reached criterion level within the minimum of two blocks during Experiment 3, selecting the functional rake significantly above chance levels; however, gibbons required an extended learning period in order to display this performance. This may suggest lower levels of understanding of the task demands in gibbons; however, given the potential difficulty in visually discriminating between the rakes due to the angle which most subjects approached the apparatus from, further research is needed in order to assess gibbons' spontaneous understanding during this task.

That subjects now succeeded with this rake set suggests their learning may have been 'scaffolded' by experience of differentiating between rakes with more salient perceptual differences during Experiments 1 and 2. Scaffolding learning refers to the process by which the elements of a task which are initially beyond an individual's capabilities are controlled, permitting them to focus on the elements which they are competent with, gradually increasing the task difficulty until a successful outcome is reached (Wood, Bruner & Ross, 1976). Although scaffolding learning often incorporates a social element, with a 'teacher' aiding skill development, physical artefacts associated with tasks may also aid in the development of expertise. Considering tool use in non-human primates, Fragaszy et al. (2013) suggest artefacts, such as previously used tools or partially processed food items as a result of tool use, may guide practice in the absence of other individuals. Skilful object manipulation requires practice, and object play shows a consistent relationship with behaviours such as innovation rate, rates of extractive foraging, and tool use (Fragaszy et al., 2013; Montgomery, 2014). Thus, subjects may have built upon knowledge gained through successfully manipulating the rakes during Experiments 1 and 2, transferring this experience across rake sets in order to now succeed with the rake set presented during Experiment 3.

Experiment 4 aimed to assess whether subjects' selections were tied to the shape of the rakes presented across Experiments 1-3, with subjects selecting the rectangle-shaped functional rake regardless of what it was paired with. Experiment 4 therefore presented a different shaped functional rake, with most subjects again reaching criterion level with this rake set, suggesting they were potentially transferring knowledge of functionality learned across rake sets. As the same non-functional rake from Experiments 1 and 2 was presented during Experiment 4, we cannot rule out the possibility that successful subjects may have attended to the shape of this rake, rather than the functionality of the novel functional rake. Truman failed to reach criterion during this experiment, despite doing so in the minimum of 20 trials across previous experiments. It is possible this subject's learning was tied to the perceptual features of the functional rake presented during the repeat of the pilot testing and Experiment 2, rather than attending to both the functional and non-functional rakes. Thus, when the shape of the functional rake was changed during Experiment 4, he failed to relearn which aspects of the rakes to attend to within the limit of 30 trials.

The majority of subjects significantly selected the functional rake across Experiments 3 and 4. Performance levels were again lower during the initial trials of each experiment than the final trials of the previous experiment, suggesting subjects may have been re-learning which aspects of the rakes to attend to. However, given the high performance levels, and that subjects generally reached criterion level within a smaller number of blocks as experiments progressed, the findings suggest gibbons were generalising knowledge across rake sets. These findings again support those of Fujita et al. (2003), who found capuchins (*C. apella*) transferred knowledge across rake sets of different shapes, suggesting subjects learned something other than a specific response to a specific rake shape. Capuchins received 24 trials in each experiment presenting a change in rake shapes, selecting the functional rake significantly across these trials. These findings contrast with the drop in performance levels displayed by some gibbons during the initial trials of each experiment. It is possible the increased tool competency displayed by capuchins (see Shumaker et al., 2011, for a review) resulted in faster attentiveness to functionally relevant features.

Assessments of the ability to distinguish between functionally relevant and irrelevant features during raking in tasks with New World monkeys have frequently presented tasks were the reward position, placed either in front of, or to one side of the rake, may have been used to guide selections (Cummins-Sebree & Fragaszy, 2005; Fujita et al., 2003; Hauser, 1997; Hauser et al., 2002;

Spaulding & Hauser, 2005). Position of the reward relative to the rake could have explained performance in Experiments 1, 2 and 4. However, this could not have been used to guide selections during Experiment 3, where gibbons were making correct selections when the reward was positioned in front of both rakes. Povinelli and Reaux (2000) considered two potential models in accounting for chimpanzees (*P. troglodytes*) spontaneous understanding during an inverted rake task with similar apparatus to that presented during Experiment 3 (see Chapter 2, section 2.2 for more details). The 'perceptual containment model' was proposed to account for naïve chimpanzees initial responses on this task, with consideration of whether the reward was positioned in front of the rake or not. In contrast, the 'physical contact model' requires subjects to display an understanding of whether the rake will come in to contact with the reward or not. The increased difficulty of tasks requiring consideration of whether a rake will come into contact with a reward is supported by findings of Santos et al. (2006), who found vervet monkeys (*C. aethiops*) and tamarins (*S. oedipus*) displayed lower levels of performance during a replication of the inverted rake task than during a task where the rewards positioning could be used to guide selections. The perceptual containment model may account for gibbons' performance during Experiment 1, 2 and 4; however, this model cannot account for subjects' performance during Experiment 3. It is possible subjects considered whether the rake would come into contact with the reward; however, given that the same functional rake was used across Experiments 1-3, subjects may have been attending to the shape of this rake, rather than considering physical contact between the rakes and rewards.

The small number of both *Symphalangus* and *Nomascus* subjects may have obscured genera differences; however, there was some consistency in performance between the genera, in terms of frequency of functional rake selections. During Experiment 1, *Nomascus* subjects reached criterion level in the smallest number of blocks, suggesting these subjects learned which aspects of the rakes to attend to fastest, followed by *Hoolock*, *Hylobates* and finally, the single *Symphalangus* subject. *Hoolock* subjects reached criterion within fewer blocks than *Hylobates* or *Nomascus* subjects across Experiments 2 and 3. Furthermore, all of the *Hoolock* subjects reached criterion level in Experiment 2, suggesting either their selections during Experiment 1 had not been tied to the more salient colour differences between the rakes, or these subjects displayed a faster processing time in learning which features of the rake to attend to. In contrast, three *Hylobates* subjects failed to reach criterion level, and the *Nomascus* subjects took longer to reach criterion level in Experiment 2, suggesting some of these subjects selections had been tied to the colour differences between the rakes during Experiment 1. The trend for increased performance in *Hoolock* subjects supports previous research finding these subjects consistently outperformed other genera (*Nomascus*, *Hylobates* and *Symphalangus* species) during tool use and object manipulation tasks (Cunningham, 2006). However, during Experiment 4, *Hoolock* subjects took longest to reach criterion level. The change in the functional rake shape during this experiment suggests *Hoolock* subjects were attending to the shape of this rake across previous experiments, with the novel functional rake presented in Experiment 4 resulting in the lower performance during this experiment, as these subjects re-learned which aspects of the presented rakes to attend to.

Symphalangus and *Hylobates* subjects showed a more varied engagement with the task than the other genera. The single *Symphalangus* subject was fastest to first productive contact and spent less time attending to the task during Experiment 1 and 3, and *Hylobates* subjects the least time during Experiments 1 and 4. Given that there was only one *Symphalangus*, conclusions about the effectiveness of her engagements are tentative. However, *Hylobates* and *Symphalangus* subjects' performance levels did not appear to be related to the amount of time they spent engaged with the task, with these subjects not displaying lower frequencies of functional rake selections during those experiments where they displayed faster behaviours.

Hoolock subjects showed a general trend for spending less time engaged with the task. As these subjects also reached criterion level in fewer blocks than *Hylobates* or *Nomascus* subjects during Experiments 2 and 3, it is possible these subjects learned which aspects of the rakes to attend to faster, and successfully transferred this knowledge across experiments. *Nomascus* subjects displayed a different approach to the task, consistently spending the longest time attending to the task and being slower to first productive contact than either *Hoolock* or *Hylobates* subjects. This increased time spent considering the task did not reliably lead to an improved performance by

Nomascus subjects, who displayed a varying performance in terms of functional rake selections across the experiments.

Differences in ecology may account for the difference in approach to the task by *Hoolock* and *Nomascus* subjects. These species are found in more northern locations than other genera, resulting in a greater seasonal variation in both food availability and weather conditions (Brandon-Jones et al., 2004; Fan, Ai, Fei, Zhang & Yuan, 2013). This increased environmental variation may result in increased investigative behaviour and responsiveness to resources. The current findings suggest *Hoolock* subjects displayed more efficient behaviours, being more responsive to the changes between the rakes and learning to select the functional rake in less time than other genera. *Nomascus* subjects did display an increase in investigative behaviour; however, this did not lead to an improved performance. Given the small number of *Nomascus* subjects tested (n = 2), further research is needed with a greater number of *Nomascus* individuals in order to assess whether this increased time spent engaged with tasks leads to improved task performance within this species.

Across all experiments adults showed a general trend for performing better and spending longer engaged with the task than sub-adults. These findings are similar to those seen during pilot testing (see Chapter 3, section 3.3). During pilot testing, age was categorised as adult, sub-adult or adolescent (as classified by Palombit, 1994). As only one subject was within the adolescent age range in the current experiments, this subject was grouped with sub-adults, with subjects therefore categorised as sub-adult or adult. Although the sub-adult category in these experiments ranged from 4-8 years (compared to 6-8 years in Chapter 3), performance of the sub-adult groups was consistent with previous results in that these subjects spent less time engaged with the task and selected the functional rake less frequently than adults.

The improved performance of adult subjects is consistent with previous reviews finding higher levels of innovative behaviour and problem solving in older primates (Griffin & Guez, 2014; Reader & Laland, 2001). Innovation may build upon other skills, with the increased experience and object manipulation competency in adult individuals potentially leading to more goal directed manipulations and increased attentiveness to functionally relevant features of objects. In line with Griffin and Guez's (2014) conclusions, adults spent more time in reward-directed manipulations, suggesting more focussed, goal-directed, behaviour. However, the generally increased attentiveness to the task and latency to first productive contact by adults contrasts with previous findings. Kendal et al. (2005) found attentiveness to an extractive foraging task decreased in *Callitrichid* monkeys as age increased, with adults being faster to first successful manipulation with the task. It is suggested that differences in the task type may have resulted in the increased time adults spent engaged with the task in the current experiments. There are no existing reports of wild gibbons using objects to retrieve out of reach food sources (Shumaker et al., 2011), and as such the present task may have required less natural behaviour than that displayed during the foraging task of Kendal et al. (2005), resulting in an increased cognitive demand. Given their life histories, the increased experience of foraging behaviour may have resulted in faster successful manipulations and lower interest levels displayed by adult monkeys. In contrast, the raking in task may have presented a more difficult challenge, with adult gibbons displaying higher levels of performance after an increased time spent considering the task.

Females, who generally displayed higher frequencies of functional rake selections than males, showed a trend for spending longer attending to the task and being slower to first productive rake contact. The slightly increased latencies to interact with the task displayed by females supports previous research in suggesting females may be more risk aversive than males, displaying more cautious behaviour in response to novel objects (Cunningham, Anderson & Mootnick, 2011; Reader & Laland, 2001). However, the finding that females were generally more successful at reward retrieval contrasts with Reader and Laland's (2001) review of innovation in primates that suggested males may exhibit greater levels of innovative behaviour that would presumably be due in part, to greater experiences in successfully using objects to achieve goals. Given the generally larger body size of male primates (Bean, 1999), increased innovative behaviour and responsiveness to novel objects may confer a foraging advantage related to their increased nutritional needs (Reader & Laland, 2001). The general lack of sexual dimorphism found within gibbons (Frisch, 1963; Reichard, Barelli, Hirai & Nowak, 2016), may result in this increased responsiveness to potential food sources being less pronounced, with the better performance displayed by female subjects due to their more cautious behaviour leading to longer considering the task before selections.

That increased consideration of the task before selection led to increased performance is supported by the finding that successful individuals, those who reached criterion level performance, generally spent longer attending to the apparatus across Experiments 2-4 (all subjects were successful during Experiment 1), and were slightly slower to interact with the rakes across the potentially more challenging tasks in Experiments 3 and 4. Successful subjects were faster to first productive rake contact during Experiment 2. This was also the only experiment where successful and unsuccessful subjects selected the functional rake at equivalent levels, on average per block, potentially as a result of less time spent considering the task before selections.

In contrast to the findings of Chapter 3 (see section 3.3), the current findings support Spaulding and Hauser's (2005) hypothesis that typically non-tool using species may possess an innate mechanism for attending to the functionally relevant features of objects, with task relevant experience needed in order for successful discriminations to be made. There was some evidence of subjects re-learning which aspects of the rakes to attend to during the initial trials of each experiment. However, differences in performance levels and time spent engaged with the task between the final trials of each experiment and the initial trials of the following experiment decreased as testing progressed, suggesting subjects were learning which aspects to attend to faster, and transferring some knowledge between rake sets. Given that the same rakes were presented across different experiments, it is possible gibbons transferred knowledge of specific shape differences, rather than attending to differences in functionality between the rakes. It is also unclear from the current findings whether subjects displayed an understanding of physical contact between the rakes and reward during Experiment 3, or their selections were guided by the functional rake shape. However, the current experiments suggest some gibbons can attend to the relevant features of rakes for reward retrieval when both rakes and rewards are present during rake selections. Given this ability, Chapter 5 assessed whether gibbons could select a functional rake when the reward was absent during the rake selection period. This task required connection of temporally and spatially distinct events of rake selection and reward retrieval, assessing if gibbons could select an appropriate rake for future use.

Chapter 5: Selecting functional rakes for future use

5.1 Introduction

The findings of Chapter 4 suggest some gibbons (*Hylobatidae*) are capable of attending to the functionally relevant features of rakes when both the rewards and rakes are present during the selection period. This chapter aimed to build on those findings, assessing whether gibbons could continue to select a rake functional for reward retrieval once the food reward was absent, but could be obtained at a second location. A time delay was then introduced between rake selection and reward retrieval assessing whether gibbons could display basic prospective cognition, connecting these distinct temporal and spatial events in order to select a functional rake for future use.

Many species demonstrate future oriented behaviour, ranging from relatively inflexible, instinctual behaviour such as fixed action patterns, to episodic future thinking whereby an individual can project their mental self into the future, constructing mental simulations of potential future situations. There is debate over the criteria for assessing future thinking in non-human primates; with some suggesting that episodic future thinking is a uniquely human trait (see Suddendorf & Corballis, 2007, for a discussion of the human uniqueness of foresight). Suddendorf and Corballis (2007, 2010) suggested four criteria for assessing episodic future thinking in non-human animals. This included the use of single trials and novel problems to minimise associative learning and innate responses respectively. Furthermore, a distinct temporal-spatial separation should be imposed between action and consequence in order to avoid behavioural cueing. Finally, tasks should incorporate a range of domains, in line with humans' ability to flexibly display future oriented behaviour across domains such as technical and physical aspects of the world and social domains. In addition to these criteria, the 'Bischof-Köhler' hypothesis has been proposed (see Suddendorf & Corballis, 1997) suggesting non-human animals are incapable of anticipating future needs different from those currently experienced.

Tasks assessing planning in social domains may present an increased cognitive complexity over those that require planning within tool using domains. Social tasks, such as token exchange paradigms, may require subjects to plan what to give, where, when, and to whom, in combination with requiring cooperation with either an experimenter or another individual (Dufour & Sterck, 2008). Assessment of the ability to plan for the future within the domain of tool use removes consideration of these dynamic social dimensions, whilst still requiring subjects to act in the present moment in order to secure a future reward. Planned tool use studies have provided some support for future thinking in large apes, although the findings are criticised (see Thom & Clayton, 2016, for a review of current research, a more detailed discussion of this research is presented in Chapter 2, section 2.3).

Mulcahy and Call (2006) assessed future planning in large apes (*Pan paniscus*, n =5, *Pongo pygmaeus*, n = 5), finding subjects would select and save functional tools for future use. During Experiments 1 and 2, subjects were presented with a range of two functional and six non-functional tools in the presence of a visible (but inaccessible) apparatus containing a food reward. Subjects were required to select a functional tool, transport this tool to a waiting room, and then return with it in order to retrieve the food reward. Apes were successful with this task both when a delay period of one hour was imposed between tool selection and reward retrieval (Experiment 1) and when an overnight delay of 14 hours was imposed (Experiment 2). Subjects continued to select a functional tool during Experiment 3, where the apparatus was hidden from view during the selection, controlling for the possibility of visual access to the task cueing subjects selections.

Despite Mulcahy and Call (2006) concluding these apes displayed future planning, these findings have been challenged (see Thom & Clayton, 2016; Suddendorf & Corballis, 2010) as they did not control for subjects' motivational state during tool selection. Thus, in line with the Bischof-Köhler hypothesis, although subjects may have anticipated a future need for the tool, they may not have anticipated a future motivational state. Osvath and Osvath (2008) sought to address these criticisms during a series of experiments with large apes (*Pan troglodytes*, n = 2, *Pongo abelii*, n = 1). A self-control task was given, presenting a choice between four objects, one of which was a functional tool useable to retrieve a large reward at a later time, and an immediately available small reward. This task set current desires, for the small reward, in opposition with future desires, for the large reward. Apes preferentially selected the functional tool, suggesting subjects were capable of differentiating between current and future needs, displaying self-control in order to select the functional tool to satisfy a future desire for the larger reward.

Further criticism of Mulcahy and Call's (2006) findings suggested apes' behaviour may be attributed to associative learning, with some suggesting subjects may have formed associations between the tool and reward rather than selections being based on an anticipation of future use (see Suddendorf & Corballis, 2010; Thom & Clayton, 2016). However, during an association control task, Osvath and Osvath (2008) found subjects behaved as though they saw a tool as an instrument for future reward retrieval. Subjects were given two selection periods; the first presented a choice between one functional and three non-functional tools, and the second presented the same tools, with a small food reward also presented. Following a delay period, subjects were given access to the large reward, which was retrievable if subjects had selected the functional tool during either of the selections periods. Thus, if subjects saw the tool as a means to an end, it was expected they would maximise reward retrieval, selecting the tool during the first selection period and, since two tools would not lead to retrieval of more of the large reward, the small reward during the second selection period. Conversely, if subject's selections were based on an arbitrary association between the tool and large reward, it was expected subjects would select the tool during both selection periods. Findings suggested subjects saw the tool as a means to an end, preferentially selecting the tool during the first selection period, and the small reward during the second. Osvath and Osvath (2008) note this does not suggest the apes did not form any associations between the rake and the reward, as all prospective behaviour must be cued either externally or internally by the current situation, but rather the association was not an arbitrary one expected from simple associative learning accounts.

A final experiment conducted by Osvath and Osvath (2008) investigated ape's ability to mentally represent the particularities of a future tool using task, assessing whether subjects could distinguish between novel potential tools based on the specific requirements of the future task. Subjects were presented with a range of three novel items, one of which was functional for future reward retrieval, and one familiar item that had previously been associated with food retrieval but was non-functional for the current task. Apes were found to select the novel functional tool. These findings lessen the applicability of learned stimulus-response associations in accounting for subjects' behaviour, with Osvath and Osvath (2008) concluding the findings across the experiments show support for the ability to mentally represent future events in large apes. However, Roberts and Feeney (2009a) suggest apes may have been attending to the functional properties of the items, selecting a tool which they knew had functional value for retrieval of a large reward, without anticipating its future use. This criticism was addressed by Bräuer and Call (2014), presenting large apes (*P. troglodytes*, n = 14, *P. paniscus*, n = 4, *Pongo* sp., n = 7) with a tool manufacturing period in advance of a tool use task. Subjects were presented with soft wooden boards which could be broken into strips before being given access to a tube task, requiring the use of wooden strips inserted into one end of a tube in order to push a food reward to within reach at the opposite end of the tube. Apes were found to construct functional tools for future use, suggesting they did not simply attend to the relevant functional properties of presented items without anticipating their future use, as suggested by Roberts and Feeney (2009a).

In contrast to large apes, assessment of planned tool use in monkeys (*Macaca fascicularis*, n = 6) found subjects initially failed to transport functional tools for future use (Dekleva, van den Berg, Spruijt & Sterck, 2012). Following training on a raking in task, subjects were required to select a functional rake, transport this rake to a waiting area for a time delay, and then transport the rake to a test compartment using it to rake in an out of reach reward placed on a test table. Dekleva et al. (2012) suggest subjects may have struggled to connect the temporally separate events of rake selection and reward retrieval. However, a further two experiments found subjects failed to transport functional tools when the space between rake selection and reward retrieval was reduced, with the test table visible from the selection area, potentially acting as a behavioural cue. Subjects also failed to transport rakes when the delay period between rake selection and reward retrieval was removed. These findings provided little support for future planning capacities within monkeys; however, Dekleva et al. (2012) suggest subjects' lack of interest in the rake outside of the original training context may have interfered with subjects' ability to transport rakes. Subjects were found to have a context specific preference for the rake, interacting with the tool faster, and for longer, when it was presented on a platform and baited with a food reward, as it had been during initial training with the raking in task.

A series of progressive training steps were therefore given, training subjects to transport the rakes in incremental rewarded steps. The first of these steps required subjects to manipulate a rake onto the test table in order to rake in a reward. The second stage required subjects to transport a rake from the waiting to the test table, where they would receive a reward for returning with the rake. Finally, subjects were required to directly transport the rake from the selection area to the test compartment, using the rake to retrieve a food reward. Subjects successfully learned the required behavioural sequence after experiencing these incremental steps. A delay period was then re-introduced between rake selection and reward retrieval, assessing subjects' ability to keep in mind the association between the rake and reward over a time delay. Subjects now transported functional rakes, with one subject doing so up to a delay period of 20 minutes.

Raby and Clayton (2009) proposed a framework for assessing the possession of prospective cognition in non-human animals, differentiating between behaviours that do not require a 'sense of the future', such as fixed action patterns and learnt associations, from those behaviours that do, such as prospective memory, semantic and episodic future thinking. Although not necessarily requiring the ability to mentally project ones subjective self into the future, learned associations may still be considered a form of prospection, where subjects may guide future actions by learning about the behaviours that lead to this action (Suddendorf, Bulley & Miloyan, 2018). Dekleva et al. (2012) suggest the ability to apply the trained behaviour over time delays, and the lack of an immediate reward following the target behaviour (rake selection), makes the monkeys behaviour difficult to achieve by learned associations alone. However, as subjects were trained to solve the task and as such, their transport of the rakes was likely instigated by an associatively learned connection to the future reward retrieval event, their behaviour cannot be defined as requiring a sense of the future.

Despite dispute over the level of prospective cognition displayed by large apes, the level of prospection shown by these subjects is greater than that displayed by monkeys, being able to select and construct tools for future use without specific training. Assessment of prospective cognition in non-human animals has often focussed on restrictive criteria and episodic future thinking. Investigating a broader range of future oriented behaviours allows for consideration of different levels of prospective cognition found between species, giving a clearer view of the emergence of these capacities in non-human animals and of species-specific adaptations that may confer fitness advantages specific to those species' ecological demands.

Gibbons previously displayed the ability to attend to the functionally relevant features of rakes for reward retrieval during a raking in task, potentially based on learned associations (see Chapter 4). The current research aimed to asses if gibbons (*Hylobatidae*), taxonomically apes, were capable of selecting an un-baited functional rake for future use at a second location during a raking in task. This task is not an assessment of prospective cognition requiring a sense of the future, as in episodic or semantic prospection, but assesses if gibbons are capable of maintaining a learned association between a functional rake and a food reward when a time delay is imposed between these events. Although the current subjects had previous experience with the current raking in task, they were not given step-wise training in order to relate the now separate events of rake selection and reward retrieval, as required by monkeys (Dekleva et al., 2012). It was expected, given gibbons' phylogenetic position, they would display the ability to connect these temporally and spatially distinct events of rake selection and reward retrieval.

5.2 Training

5.2.1 Subjects

Subjects were 12 gibbons housed at the Gibbon Conservation Center (see section 3.2.1), including all subjects that reached criterion level across the series of experiments presented in Chapter 4. This included five eastern hoolock (*Hoolock. leuconedys*), three northern white cheeked (*Nomascus leucogenys*), three pileated (*Hylobates pileatus*), and one siamang gibbon (*Symphalangus syndactylus*) (see Table 5.1). All subjects housed in groups or pairs were separated and tested individually, with the exception of Iszie and Ricky. Although separation would have been preferable, this was balanced with welfare issues, with these two individuals becoming distressed when separated. Iszie and Ricky were therefore tested in their family groups; however, no other individuals in their enclosures participated in testing. Iszie's group consisted of an adult male and female pair (her parents) and her infant sister. Ricky's group consisted of her male partner and their three offspring. Despite being separated for testing, in enclosures 6, 9, 10 and 15 where two individuals participated in testing, subjects may have witnessed other individuals participating in the task, potentially influencing performance. Order effects were therefore analysed, with order of testing not found to affect performance levels in any group (see relevant results sections for order effect analyses).

Table 5.1: Subjects' species, age, sex, housing information, and whether they reached criterion level performance in the training phase or not.

Subject	Enclosure	Species	Age	Sex	Group	Reached criterion
Chan Thar	6	*H. leuconedys*	9	F	Male/female pair	✓
Win Bo	6	*H. leuconedys*	12	M	Male/female pair	✓
Betty	10	*H. leuconedys*	16	F	Female & male/male siblings	✓
Khin Maung Win	10	*H. leuconedys*	8	M	Female & male/male siblings	X
U Myint Swe	10	*H. leuconedys*	6	M	Female & male/male siblings	✓
Iszie*	5	*H. pileatus*	4	F	Family	✓
Truman	9	*H. pileatus*	12	M	Solitary	✓
Violet	9	*H. pileatus*	6	F	Solitary	✓
Ricky*	13	*N. leucogenys*	30	F	Family	✓
Asteriks	15	*N. leucogenys*	16	F	Family	✓
Lucia	15	*N. leucogenys*	6	F	Family	✓
Marlow	14	*S. syndactylus*	10	F	Solitary	✓

✓ = Reached criterion level

X = Did not reach criterion level

* Not separated for testing

5.2.2 Experimental procedure and apparatus

In order for subjects to learn that a selected un-baited rake would later be presented with a food reward, they were first given a training stage with one functional (F1, as described in Table 4.2) and one non-functional rake (NF2, as described in Table 4.2), placed on a raised platform (as described in section 3.2.2) outside their enclosure. Subjects were required to initially select a rake in the absence of any reward, with the rake baited following selection if the functional rake was chosen (see Figure 5.1). Although the rakes were presented un-baited, it was possible subjects had retained some knowledge of rake functionality from experience with this rake set in previous experiments (see Chapter 4). If not, it was expected subjects would learn to reliably select the functional rake through attending to the functionally relevant features over testing as was found during previous experiments manipulating perceptual features of presented rakes (see Chapter 4). A five second delay took place after setting of the apparatus to try and minimise impulsive choices and encourage consideration of the task, before the rakes were pushed to a starting position within subjects' reach.

As across previous experiments (see Chapters 3 and 4), subjects were considered as making a selection once one of the rakes had been pulled towards them by 10cm or more. If the functional rake was selected, the non-functional was removed from subjects' reach, with the functional pulled back to the starting position and baited immediately. Subjects were then able to pull in the rake and retrieve the reward. The immediate baiting of the functional rake following selection aimed to encourage subjects to learn the association between the functional rake and reward, before an increased time delay between rake selection and reward retrieval was introduced in later experiments (see section 5.3). The food reward used throughout testing was blueberries. If the non-functional rake was selected, both rakes were removed from reach and the trial ended. Successful reward retrieval therefore depended on the selection of the functional rake, with subjects unrewarded if the non-functional was chosen.

Figure 5.1: Northern white cheeked gibbon (*N. leucogenys*) selecting the un-baited functional rake during training, the non-functional rake remains in the starting position.

The position of the functional rake was randomised, appearing on both sides equally, but never on the same side more than twice in a row. Trial duration was a maximum of 30 minutes; if a selection was not made within this time, testing was discontinued with this subject, continuing the following day. Trials were presented in blocks of ten, with subjects given a maximum of two blocks per day. Inter-trial intervals of 20 seconds and inter-block intervals of five minutes were enforced. Subjects were required to select the functional rake 8/10 times on two consecutive blocks to be considered as reaching criterion, with a maximum of 12 blocks (120 trials) presented to each individual. All trials were recorded using a video camera for later behavioural coding. Functional rake position (left/right) and rake selected (functional/non-functional) were noted during testing.

5.2.3 Behavioural variables

As subjects were free to move around their enclosures during testing, an area 1m^2 surrounding the testing apparatus was designated as the target area. The time taken for a subject to first enter the target area was coded, along with the time taken to first productive rake contact (see Table 5.2). These variables measured latency to engage with the task and cognitive processing time, by assessing how long subjects took after entering the testing area before contact with the rakes. Time to first productive rake contact was taken from the time when subjects entered the target area and attended to the apparatus, until the first rake contact that moved the rake towards the subject. Behavioural coding was completed using 'Solomon Coder', a video coding program available online (Péter, 2018). As in Chapter 4, age was categorised as adult (8+ years), or sub-adult (4-8 years).

Table 5.2: Definitions of the behaviours coded from trial videos.

Behaviour	Description
Time to first enter target area	Time taken from the start of the trial to when subjects first entered the designated target area.
Time to first productive rake contact	Time to the first contact that moved the rake towards the subject in a reward directed manner. Taken from the time when subjects entered the target area and attended to the apparatus until first reward directed contact.

5.2.4 Result and discussion

5.2.4.1 General performance overview

Of the 12 subjects, 11 reached criterion level performance (selection of the functional rake on 8/10 trials over two consecutive blocks), taking between two and five blocks to do so (see Table 5.3). Performance across all trials completed by each subject was evaluated using binomials tests, with all subjects other than Khin Maung Win performing significantly above chance levels. Khin Maung Win also failed to reach criterion level and so was excluded from further testing. Findings therefore suggest gibbons were able to select a rake functional for reward retrieval when the reward was not in direct association with the rakes at time of choice. As with Dekleva et al.'s (2012) finding that monkeys (*M. fascicularis*) only successfully transported functional rakes for future use following progressive training experience, it is possible gibbons previous experience with the presented rakes (see Chapters 3 & 4) meant selections were guided by associatively learned relations between the functional rake and reward. The present findings therefore do not suggest gibbons acted with a sense of the future in mind, or mentally represented reward retrieval when making selections (Raby & Clayton, 2009).

Dekleva et al. (2012) suggested monkeys' failure to transport the rakes may have been due to a context specific preference for the rake within the original training paradigm when the rake was placed on a testing platform next to food rewards. Monkeys failed to relate rake selections to reward retrieval outwith this context, failing to transport rakes without step-wise training. In contrast, gibbons did not require progressive steps in order to relate rake selection and reward retrieval events. Six subjects displayed a spontaneous understanding of the task, reaching criterion within the minimum of two blocks. The remaining five successful subjects requiring a period of learning before displaying criterion level performance.

Table 5.3: Subjects' genus, age group and sex. Number of correct selections of the functional rake in each block completed by subjects during training (1 block = 10 trials). Blocks where subjects reached criterion level performance are highlighted. Significance on binomial tests for number of correct selections of the functional rake across the total number of trials completed by each subject (Total number of correct trials), and during blocks where criterion level performance was achieved (Performance on criterion blocks) are also shown.

Subject	Genus	Age group	Sex	Total number of correct trials per block												Total number of correct trials	Performance on criterion blocks
				1	2	3	4	5	6	7	8	9	10	11	12		
Betty	*Hoolock*	Adult	F	9	8											17/20**	17/20**
Chan Thar	*Hoolock*	Adult	F	8	5	9	9									31/40**	18/20**
Khin Maung Win	*Hoolock*	Adult	M	7	4	8	6	4	3	6	7	4	3	6	6	64/120	n/a
U Myint Swe	*Hoolock*	Sub-adult	M	9	7	8	10									34/40**	18/20**
Win Bo	*Hoolock*	Adult	M	8	5	9	10									32/40**	19/20**

Iszie	*Hylobates*	Sub-adult	F	10	8											18/20**	18/20**
Truman	*Hylobates*	Adult	M	9	8											17/20**	17/20**
Violet	*Hylobates*	Sub-adult	F	10	10											20/20**	20/20**
Asteriks	*Nomascus*	Adult	F	8	7	7	9	10								41/50**	19/20**
Lucia	*Nomascus*	Sub-adult	F	8	8											16/20*	16/20*
Ricky	*Nomascus*	Adult	F	9	7	8	10									34/40**	18/20**
Marlow	*Symphalangus*	Adult	F	10	9											19/20**	19/20**

= Side bias

= Criterion level performance

* Significant at $p < 0.05$ level

** Significant at $p < 0.01$ level

Generalized Estimating Equations (GEE), an extension of generalized linear models for evaluating repeated measures, were used to build a model assessing the effect of age and sex on the number of correct selections of the functional rake when number of blocks completed overall was controlled for (subjects completed different numbers of blocks dependent on how many were needed to reach criterion performance) (see Table 5.4). The model was specified as Poisson loglinear with log link function, with number of correct selections of the functional rake as the dependent variable. Age group and sex were set as factors, with total number of blocks completed set as a covariate. For parameter estimates, sub-adult was set as the reference category when evaluating age differences, and female the reference category when evaluating sex differences. The working correlation matrix was specified as 'independent' meaning trials were assumed to be independent of each other. Age was categorised as adult (8+ years), or sub-adult (4-8 years), as in Chapter 4.

Table 5.4: GEE parameter estimates for factors used in model to compare frequency of functional rake selections between age groups and sexes, with number of blocks completed as a covariate.

Factor	**Parameter**	**β**	**SE**	**Lower CI**	**Higher CI**	**Wald Chi (χ^2)**	**df**	**Sig.**
	Number of blocks completed	-0.05	0.00	-0.05	-0.04	338.30	1	**<0.01**
Age group	Adult	-0.02	0.04	-0.10	0.06	0.27	1	0.60
	Sub-adult	0[a]						
Sex	Male	-0.03	0.03	-0.08	0.03	0.72	1	0.40
	Female	0[a]						

Significant results are in bold

[a] Set to zero because this parameter is redundant

Neither age nor sex were significant predictors of performance (see Table 5.4). In contrast to findings from previous chapters, sub-adults selected the functional rake marginally more frequently than adults. Females displayed a slightly higher frequency of functional rake selections than males, consistent with findings from previous experiments. The number of blocks completed significantly predicted performance ($p < 0.01$) (see Table 5.4). As expected, those subjects that reached criterion level in fewer blocks selected the functional rake more frequently per block than those who required more blocks, suggesting some subjects were learning the correct response through experience, as opposed to other individuals who displayed a more spontaneous understanding of the task, potentially based on previous learning. Genus was excluded from the GEE model as there was only one *Symphalangus* subject. The one *Symphalangus* subject and *Hylobates* subjects took 2 blocks to reach criterion level, suggesting an immediate understanding of the task demands in these subjects. Of the *Hoolock* subjects who did reach criterion, excluding Khin Maung Win, they did so in an average of 3.5 blocks. *Nomascus* subjects performed poorest, taking an average of 3.7 blocks to reach criterion level, suggesting increased difficulty in relating the un-baited rakes to the future reward in these subjects.

5.2.4.2 Latency to interaction

Two models were built using GEEs to assess the effect of age and sex on time to first enter the target area, and time to first productive contact with the apparatus when number of blocks completed was controlled for (see Table 5.5). Both models were as detailed above for assessment of performance levels (see section 5.2.4.1), except the models were now specified as linear with identity link function, with either time to first enter the target area, or time to first productive rake contact as the dependent variable.

Table 5.5: GEE parameter estimates for factors used in models to compare the time to first enter the target area and time to first productive rake contact between age groups and sexes, with number of blocks completed as a covariate.

Model	Factor	Parameter	B	SE	Lower CI	Higher CI	Wald Chi (χ^2)	df	Sig.
Time to first enter target area		Number of blocks completed	0.88	1.82	-2.70	4.45	0.23	1	0.63
	Age group	Adult	-16.61	23.54	-62.74	29.53	0.50	1	0.48
		Sub-adult	0[a]						
	Sex	Male	-5.81	16.56	-38.25	26.64	0.12	1	0.73
		Female	0[a]						
Time to first productive rake contact		Number of blocks completed	-0.07	0.11	-0.29	0.14	0.47	1	0.50
	Age group	Adult	2.31	1.05	0.25	4.37	4.81	1	**<0.05**
		Sub-adult	0[a]						
	Sex	Male	-1.78	1.00	-3.74	0.18	3.16	1	0.08
		Female	0[a]						

Significant results are in bold

[a] Set to zero because this parameter is redundant

There was no significant difference in time to enter the target area between age groups, with sub-adults being only marginally slower than adults. Sub-adults were, however, significantly faster to first productive contact with the rakes than adults ($p < 0.05$) (see Table 5.5). In contrast to previous experiments, these findings suggest longer processing time before making selections did not lead to improved performance in adults. These findings contrast with those of Kendal et al. (2005), who found adult *Callitrichid* monkeys had significantly shorter latencies to first successful manipulations during an extractive foraging task than sub-adults, suggested to be a consequence of lack of engagement. Sub-adults may have been attending to more salient perceptual differences between the rakes, such as rake-end shape, allowing for faster selections than in adults, who may have taken the rewards position in to account during previous experiments, being required to re-learn which aspects of the rake to attend to. Sex was not a significant predictor across either model (see Table 5.5). However, as found across previous experiments, females were generally slower than males to enter the target area and to make their first productive manipulation. The number of blocks completed was also not a significant predictor across either of the models (see Table 5.5).

Patterns of interaction between the genera were similar to those found across previous chapters. *Nomascus* subjects, who required the greatest number of blocks before reaching criterion, were the fastest to enter the target area, but took longest to first productive rake contact, in line with these subject's longer latencies found in previous experiments. *Hoolock* subjects, who showed a general trend for faster behaviour across previous experiments, displayed the same tendency here; however, this did not lead to improved performance in these subjects who took the greatest number of blocks to reach criterion level. *Hylobates* and the single *Symphalangus* subject, who reached criterion level within the minimum of two blocks, were both slower to approach the task than either *Nomascus* or *Hoolock* subjects. These genera displayed different approaches to the task once in the target area, with *Hylobates* being the fastest to first productive contact, and *Symphalangus* faster than only *Nomascus* subjects; however, due to there only being one *Symphalangus* subject, differences in approach style may not be representative of genera differences.

5.2.4.3 Order effects

Order effects were assessed using Mann-Whitney U tests for those individuals who may have witnessed other subjects completing testing, in enclosures 6, 9, 10 and 15, with time to first productive rake contact over the first ten trials compared. No significant difference was found for Win Bo and Chan Thar in enclosure 6 ($U = 27.5$, $z = -1.70$, $p = 0.09$). A significant difference was found between Violet and Truman in enclosure 9 ($U = 18$, $z = -2.42$, $p = 0.02$). Truman, the second individual to be tested, was faster to first productive contact with the rakes (mean = 2.21 seconds, Violet mean = 5.23 seconds) however, he selected the functional rake less frequently (9/10 trials) than Violet (10/10 trials), suggesting testing order did not aid performance. A significant difference was also found between Asteriks and Lucia in enclosure 15 ($U = 21$, $z = -2.2$, $p = 0.03$), with Lucia (mean = 1.89 seconds), the second to be tested, making first productive contact with the rakes sooner than Asteriks (mean = 2.99 seconds). However, both individuals selected the functional rake at the same frequency (8/10 trials). Although witnessing Asteriks participating in testing may have led to Lucia interacting with the task faster, testing order again did not improve task performance.

In enclosure 10, the first individual tested (U Myint Swe) was fastest to first productive contact and selected the functional rake on 9/10 trials. Betty, the second tested, was slowest to first productive contact (mean = 13.21 seconds), selecting the functional rake at the same level as U Myint Swe (9/10 trials). Khin Maung Win, the last to be tested, was faster than Betty to first productive contact (mean = 6.22 seconds), but selected the functional rake less frequently (7/10 trials). The differences between U Myint Swe and Betty ($U = 8$, $z = -3.18$, $p < 0.01$) and between Betty and Khin Maung Win were significant ($U = 20$, $z = -2.27$, $p = 0.02$). There was no significant difference between U Myint Swe and Khin Maung Win ($U = 40$, $z = -0.76$, $p = 0.48$). Testing order therefore did not appear to influence overall performance, with Khin Maun Win, the last to be tested, failing to reach criterion level suggesting he did not benefit from witnessing other individuals participating in testing.

5.3 Experiments 1-4: Selecting a functional rake for future use at a second location

Subjects who passed the training phase were then presented with a series of four experiments introducing a time delay between rake selection and reward retrieval. Furthermore, rake selection and reward retrieval now occurred at two separate testing platforms, assessing if subjects could reliably select a functional rake in order to retrieve the reward at a later time and at a different location. The time delay was initially zero seconds during Experiment 1, with the selected rake immediately taken to the second testing platform, before being increased to 30 seconds in Experiment 2, two minutes in Experiment 3, and finally five minutes during Experiment 4.

5.3.1 Subjects

Eleven subjects reached criterion level during training and passed to this stage (see Table 5.6). Marlow developed a lack of interest in the task during Experiment 1, only completing 12 trials, and was therefore excluded from further testing. The remaining ten subjects completed all delay periods (Experiment 1 – 4). Subjects were again separated and tested individually, with the exception of Iszie and Ricky who were tested in their family groups. In enclosures 6, 9, 10 and 15, where two individuals were tested, subjects may have witnessed other individuals participating in testing, despite being separated for testing. Order effects were therefore analysed for these enclosures, with the order of testing not found to affect performance levels for any experiment (see results sections for each experiment for analyses).

Table 5.6: Subjects' species, age, sex, housing information and details of which experiments each subject participated in. Whether subjects reached criterion level performance is marked for Experiments 1-3 (see footnotes), and whether subjects performed significantly above chance during the ten trials presented in Experiment 4.

Subject	Enclosure	Species	Age	Sex	Group	Exp. 1	Exp. 2	Exp. 3	Exp. 4
Chan Thar	6	*H. leuconedys*	9	F	Male/female pair	✓	✓	✓	S
Win Bo	6	*H. leuconedys*	12	M	Male/female pair	✓	✓	✓	S
Betty	10	*H. leuconedys*	16	F	Female & male/male siblings	✓	✓	✓	S
U Myint Swe	10	*H. leuconedys*	6	M	Female & male/male siblings	✓	✓	✓	NS
Ricky*	13	*N. leucogenys*	30	F	Family	✓	✓	✓	NS
Asteriks	15	*N. leucogenys*	16	F	Family	✓	✓	✓	S
Lucia	15	*N. leucogenys*	6	F	Family	✓	✓	✓	S
Iszie*	5	*H. pileatus*	4	F	Family	✓	✓	✓	S
Truman	9	*H. pileatus*	12	M	Solitary	✓	✓	✓	S
Violet	9	*H. pileatus*	6	F	Solitary	✓	✓	✓	S
Marlow	14	*S. syndactylus*	10	F	Solitary	X[1]			

✓ = Reached criterion level

X = Did not reach criterion level

S = Performed significantly above chance levels

NS = Did not perform significantly above chance levels

[1] Did not complete testing

* Not separated for testing

5.3.2 Experimental procedure and apparatus

Procedure and apparatus were identical to that of training (see section 5.2.2), with the following exceptions. Subjects were now required to select a rake at platform 1, before moving to platform 2 to use the selected rake (rake was transported to platform 2 by the experimenter). The platforms used were identical, except platform 2 had no middle wooden divider; as only one rake was ever presented on platform 2, the issue of the reward passing into the path of a different rake did not occur. As in training, an area $1m^2$ surrounding each platform was designated as a target area. The distance between the platforms differed depending on enclosure layout; however, both platforms were always visible to the subject from any position in the target areas. During Experiment 1, with a 0 second delay, the chosen rake was immediately taken to the second platform by the experimenter following selection, and baited, regardless of whether the functional or non-functional rake was chosen. Subjects were then able to attempt to retrieve the reward by using the chosen rake to draw the food within reach; if they had selected the functional rake, they could obtain the reward. If the non-functional rake was selected, this was also taken to the retrieval platform and baited. Although this rake was could not be used to obtain the reward, baiting of the non-functional rake aimed to reinforce subjects learning of the functional differences between rakes. Trials ended once the subject had pulled the baited rake towards themselves 10cm or further. Presentation side was consistent across platforms, with the selected rake being placed on the same side of the retrieval platform (platform 2) as it appeared at the selection platform (platform 1).

Time delays were then introduced between rake selection and use. The delay was initially 30 seconds during Experiment 2, being increased to two minutes in Experiment 3 and finally five minutes in Experiment 4. It was expected this increasing delay would increase the difficulty in relating rake selection to reward retrieval, with the incremental steps in delay time aiming to assess if subjects' performance declined at a specific delay period. In line with previous findings suggesting shorter inter-trial intervals may result in interference in learning between trials, the inter-trial interval was increased to an equivalent time as the delay period in each experiment (Jarrard & Moise, 1971). This aimed to make the separate trials more distinctive, encouraging subjects to relate the rake selection period of each trial to the delayed reward retrieval period. For example, if on trial two, rake selection occurred 20 seconds following trial one's reward retrieval period, subjects may have associated these two events, rather than linking trial one's reward retrieval period to the rake selection period which had occurred before the imposed time delay.

During Experiments 2-4 the unselected rake was removed from the testing platform, with the selected rake left within subjects' reach on platform 1 until the end of the delay period, at which time the experimenter moved it to the retrieval platform (see Figure 5.2). Due to decreasing motivation levels during trials with delay periods, the reward size was increased to three blueberries during Experiments 2-4. Subjects were permitted up to 12 blocks (120 trials) to reach criterion level during Experiments 1-3. Due to testing time constraints, the number of trials presented to each subject was limited to one block of ten trials during Experiment 4, presenting an initial assessment of gibbons' abilities at this time delay.

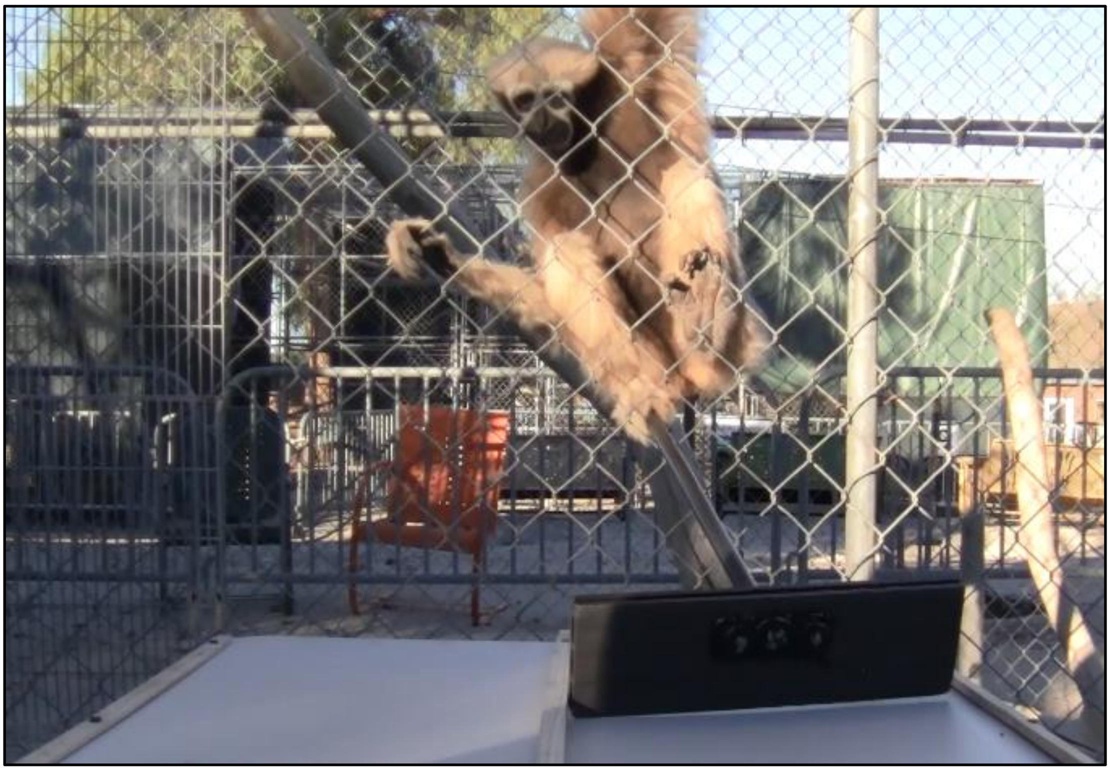


Figure 5.2: Eastern hoolock gibbon (*H. leuconedys*) interacting with the selected functional rake during a delay period.

5.3.3 Behavioural variables

The behavioural variables coded were identical to training (see section 5.3.2), with the addition of the time subjects spent in the target area surrounding platform 1 attending to the task during the delay period also recorded during Experiments 2-4 (see Table 5.7). Time to first approach at platform 1 allowed for measurement of subjects' latency to interact with the task as the delay period increased and served as an assessment of motivation. Time to first productive rake contact was also measured at platform 1, assessing subjects' cognitive processing time between entering the target area, and first contact with a rake in a manner that moved the rake towards them, as required for rake selection. This was taken from the time subjects entered the target area until first productive contact with either rake.

The selected rake was left within subjects' reach for the duration of the delay period, at platform 1. As such, time spent attending to the task at platform 1 during the delay period of Experiments 2-4 began immediately following rake selection until the specified delay time had elapsed. This allowed for assessment of whether time spent attending to the task increased as the delay period increased, suggesting subjects were returning to the testing area more frequently, potentially to keep the task in mind, as the delay became longer. Generalized Estimating Equations (GEE) were used to build four models assessing the effect of delay period, genus, age and sex on frequency of functional rake selections and each of the three behavioural variables listed in Table 5.7. The specifications and findings of these models are reported in section 5.3.8.

Table 5.7: Definitions of the behaviours coded from trial videos. Time to first approach and to first productive rake contact were recorded for Experiments 1-4. Time spent attending to the apparatus during the delay period was recorded for Experiments 2-4 only. All behaviours were recorded at platform 1.

Behaviour	Description
Time to first approach	Time taken from the start of the trial to when subjects first entered the target area surrounding platform 1.
Time to first productive rake contact	Taken from the time when subjects entered the target area surrounding platform 1 and attended to the apparatus, until first contact that moved the rake towards the subject. Subjects were required to pull the rake towards them 10cm or more to make a selection.
Time spent attending to the apparatus during the delay period	Time spent either visually attending to, or in physical contact with (through hands, feet or mouth), the apparatus at platform 1 during the delay period. The delay period began immediately following rake selection and ended once the designated delay

	period had passed (30 seconds, 2 minutes, or 5 minutes).

5.3.4 Experiment 1: Zero delay period results

All ten subjects who completed testing reached criterion level performance within four blocks in Experiment 1 (0 seconds delay) (see Table 5.8). Marlow became disinterested in the task, only completing 12 trials, and so was excluded from analyses. Performance across all trials completed by each subject was evaluated using binomials tests with all subjects performing significantly above chance levels across all trials (see Table 5.8). These findings again suggest gibbons are capable of selecting an appropriate rake for reward retrieval when the relationship between the rake and reward is not visible at the time of selection. Furthermore, the use of the rake at a second location suggests physical association between the reward and rake was not necessary for success. Gibbons displayed the ability to connect the temporally and spatially separate events of rake selection and reward retrieval within few trials. Seven subjects reached criterion within the minimum of two blocks of trials, suggesting the implementation of a separate retrieval platform did not require subjects to re-learn the association between the rakes and rewards. These findings again contrast with those of Dekleva et al. (2012) who found monkeys (*M. fascicularis*) displayed a context specific preference for rake use within the original training context, presented baited on a testing platform, failing to transfer this behaviour to a different context without step-wise training. As with training, it is possible subjects' selections were guided by an associatively learned connection between the rakes and rewards across previous experiments (see Chapters 3 and 4). These findings are therefore not considered as requiring a sense of the future or mental representation of the reward retrieval event (Raby & Clayton, 2009).

Table 5.8: Subjects' genus, age group and sex and number of correct selections of the functional rake in each block completed during Experiment 1 (1 block = 10 trials). Blocks where subjects reached criterion level performance are highlighted. Number of correct selections of the functional rake across the total number of trials completed by each subject, and during the blocks where criterion level performance was achieved are also shown, with significance values from binomial tests.

Subject	Genus	Age group	Sex	Total number of correct trials per block				Total number of correct trials	Performance on criterion blocks
				1	2	3	4		
Betty	*Hoolock*	Adult	F	10	10			20/20**	20/20**
Chan Thar	*Hoolock*	Adult	F	6	7	8	9	30/40**	17/20**
U Myint Swe	*Hoolock*	Sub-adult	M	8	8			16/20*	16/20*
Win Bo	*Hoolock*	Adult	M	9	8			17/20**	17/20**
Iszie	*Hylobates*	Sub-adult	F	6	10	10		26/30**	20/20**
Truman	*Hylobates*	Adult	M	9	9			18/20**	18/20**
Violet	*Hylobates*	Sub-adult	F	9	10			19/20**	19/20**
Asteriks	*Nomascus*	Adult	F	10	7	10	9	36/40**	19/20**
Lucia	*Nomascus*	Sub-adult	F	9	8			17/20**	17/20**
Ricky	*Nomascus*	Adult	F	9	10			19/20**	19/20**

■ = Criterion level performance

* Significant at $p < 0.05$ level

** Significant at $p < 0.01$ level

5.3.4.1 Order effects

Order effects were assessed using Mann-Whitney U tests for those individuals who may have witnessed other subjects completing testing, with time to first productive rake contact over the first ten trials compared. There was no significant difference between Win Bo and Chan Thar in enclosure 6 ($U = 32$, $z = -1.36$, $p = 0.19$). A significant difference was found in enclosure 9 between Violet and Truman ($U = 10$, $z = -3.03$, $p < 0.01$). The second individual to be tested (Violet), was slower to first productive contact (mean = 6.27 seconds, Truman mean = 2.01 seconds), with both individuals selecting the functional rake at the same level (9/10 trials). A significant difference was also found between U Myint Swe and Betty in enclosure 10. Betty, the second individual to be tested, selected the functional rake more frequently (10/10 trials) than U Myint Swe (8/10 trials) but took longer to first productive contact (mean = 10.47 seconds, U Myint Swe mean = 3.33 seconds). This suggests Betty may have benefitted from testing order; however, given that Betty has consistently displayed higher performance levels than U Myint Swe, her better performance may represent cognitive differences in these subjects. Finally, in enclosure 15, there was a significant difference between Asteriks and Lucia ($U = 21$, $z = -2.2$, $p = 0.03$). Lucia, the second subject tested, was faster to first productive contact (mean = 2.37, Asteriks mean = 4.02 seconds), but selected the functional rake less frequently (9/10 trials, Asteriks 10/10 trials). Overall, witnessing other individuals participating in testing did not appear to consistently improve performance, with all subjects selecting the functional rake above chance levels.

5.3.5 Experiment 2: 30 seconds delay period results

All ten subjects reached criterion level performance, with nine reaching criterion in the minimum of two blocks (see Table 5.9). Performance across all trials completed by each subject was evaluated using binomials tests, with all subjects performing significantly above chance levels (see Table 5.9). The introduction of a time delay between rake selection and reward retrieval therefore did not result in a decrease in the majority of subjects' performance, suggesting gibbons are capable of relating the separate events of rake selection and reward retrieval across this time delay. In contrast to the other subjects, Truman took seven blocks in order to reach criterion. It is possible this subject was required to re-learn the association between rake selection and reward retrieval once a time delay was imposed, making these events more temporally distinctive. More likely, given that Truman selected the functional rake significantly frequently across all trials, and displayed criterion level performance on block one, it is possible the introduction of a delay temporarily decreased motivation in this subject.

Table 5.9: Subjects' genus, age group and sex and number of correct selections of the functional rake in each block completed by subjects during Experiment 2 (1 block = 10 trials). Blocks where subjects reached criterion level performance are highlighted. Significance on binomial test for number of correct selections of the functional rake across the total number of trials completed by each subject is also shown.

Subject	Genus	Age group	Sex	Total number of correct trials per block							Total number of correct trials
				1	2	3	4	5	6	7	
Betty	*Hoolock*	Adult	F	10	10						20/20**
Chan Thar	*Hoolock*	Adult	F	10	10						20/20**
U Myint Swe	*Hoolock*	Sub-adult	M	8	8						16/20*
Win Bo	*Hoolock*	Adult	M	9	8						17/20**
Iszie	*Hylobates*	Sub-adult	F	9	9						18/20**
Truman	*Hylobates*	Adult	M	8	7	6	6	7	9	8	51/70** (17/20** on criterion blocks)
Violet	*Hylobates*	Sub-adult	F	9	9						18/20**
Asteriks	*Nomascus*	Adult	F	10	9						19/20**
Lucia	*Nomascus*	Sub-adult	F	9	8						17/20**
Ricky	*Nomascus*	Adult	F	8	8						16/20*

■ = Criterion level performance

* Significant at p < 0.05 level

** Significant at $p < 0.01$ level

5.3.5.1 Order effects

Mann-Whitney U tests were used to assess order effects for those individuals who may have witnessed other subjects completing testing, with time to first productive rake contact over the first ten trials compared. There was no significant difference found between Win Bo and Chan Thar ($U = 36.5$, $z = -1.02$, $p = 0.32$), U Myint Swe and Betty ($U = 48.5$, $z = -0.11$, $p = 0.91$), or Asteriks and Lucia ($U = 29$, $z = -1.6$, $p = 0.12$). A significant difference was found between Violet and Truman in enclosure 9 ($U = 15.5$, $z = -2.61$, $p = 0.01$). Violet, the second individual to be tested, selected the functional rake more frequently, but was slower to first productive contact (mean = 5.07 seconds, Truman mean = 4.2 seconds). Although this result does suggest some learning may have taken place, Truman selected the functional rake on 8/10 trials, compared to Violet's selection of the functional rake on 9/10 trials. Therefore, it is difficult to draw conclusions here that witnessing Truman participate in testing was of benefit to Violet as both subjects displayed high levels of performance during the initial ten trials. Given the lack of evidence of order effects in all previous experiments, it was therefore assumed testing order did not affect overall performance levels.

5.3.6 Experiment 3: 2 minutes delay period results

All ten subjects reached criterion level performance, with all but U Myint Swe and Ricky reaching criterion in the minimum of two blocks (see Table 5.10). Performance across all trials completed by each subject was evaluated using binomials tests, with all subjects performing significantly above chance levels (see Table 5.10). Truman, who required more blocks to reach criterion level during Experiment 2 than other subjects (see section 5.3.5) now reached criterion within the minimum of two blocks. If this subject had been re-learning the association between the rakes and reward following the implementation of a time delay during Experiment 2, he appeared to successfully transfer this learning to the longer time delay during Experiment 3. The increased number of blocks taken to reach criterion level by U Myint Swe and Ricky during Experiment 3 may likewise represent a re-learning to connect rake selection and reward retrieval events by these subjects. Overall, the increasing delay period did not appear to affect the majority of subjects' performance, suggesting gibbons continued to relate the temporally distinct events of rake selection and reward retrieval once a two minute delay was imposed.

Table 5.10: Subjects' genus, age group and sex and number of correct selections of the functional rake in each block completed by subjects during Experiment 3 (1 block = 10 trials). Blocks where subjects reached criterion level performance are highlighted. Significance on binomial tests for number of correct selections of the functional rake across the total number of trials completed by each subject are also shown.

Subject	Genus	Age group	Sex	Total number of correct trials per block								Total number of correct trials
				1	2	3	4	5	6	7	8	
Betty	*Hoolock*	Adult	F	10	9							19/20**
Chan Thar	*Hoolock*	Adult	F	8	8							16/20*
U Myint Swe	*Hoolock*	Sub-adult	M	9	7	7	8	6	5	9	9	60/80** (18/20** on criterion blocks)
Win Bo	*Hoolock*	Adult	M	9	8							17/20**
Iszie	*Hylobates*	Sub-adult	F	9	8							17/20**
Truman	*Hylobates*	Adult	M	10	9							19/20**
Violet	*Hylobates*	Sub-adult	F	10	9							19/20**
Asteriks	*Nomascus*	Adult	F	9	9							18/20**
Lucia	*Nomascus*	Sub-adult	F	10	9							19/20**
Ricky	*Nomascus*	Adult	F	7	10	8						25/30** (18/20** on criterion blocks)

■ = Criterion level performance

* Significant at $p < 0.05$ level

** Significant at $p < 0.01$ level

5.3.6.1 Order effects

Order effects were assessed using Mann-Whitney U tests for those enclosures where two individuals participated in testing, with time to first productive rake contact over the first ten trials compared. There was no significant difference found between Violet and Truman ($U = 31$, $z = -1.44$, $p = 0.17$) or Asteriks and Lucia ($U = 48$, $z = -0.15$, $p = 0.91$). A significant difference was found between Win Bo and Chan Thar in enclosure 6 ($U = 17.5$, $z = -2.46$, $p < 0.05$). Chan Thar, the second individual tested, was faster to first productive contact (mean = 2.26 seconds, Win Bo mean = 7.66 seconds) but selected the functional rake less frequently (8/10 trials) than Win Bo (9/10 trials). There was also a significant difference between U Myint Swe and Betty in enclosure 10 ($U = 20$, $z = -2.27$, $p < 0.05$). Betty, who was tested second, selected the functional rake more frequently (10/10 trials) than U Myint Swe (9/10 trials), suggesting she may have benefitted from watching U Myint Swe complete trials. However, given the small difference in performance levels between these subjects, and the finding that Betty took longer to first productive contact (mean = 8.89 seconds, U Myint Swe mean = 3.28 seconds), it appears her overall performance was not positively influenced. Testing order therefore did not appear to consistently affect performance levels.

5.3.7 Experiment 4: 5 minutes delay period results

Performance across the ten trials completed by each subject was evaluated using binomial tests, with 8/10 subjects selecting the functional rake significantly above chance levels (see Table 5.11). When compared to the first ten trials of Experiment 3, performance did not decrease during Experiment 4, with subjects selecting the functional rake an average of 9/10 trials across both experiments. These findings suggest the eight successful subjects continued to relate distinct rake selection and reward retrieval events, even across a five minute delay period.

Ricky's performance did not appear to be significantly impaired by the increasing delay period, selecting the functional rake on 8/10 trials, a comparable performance to her first ten trials during Experiment 2 (8/10 trials, see Table 5.9), and an increased frequency of functional rake selections than that shown during Experiment 3 (7/10 trials, see Table 5.10). However, given that only ten trials were presented to each subject, she did not select the functional rake significantly above chance levels during Experiment 4 (binomial test p = 0.11). U Myint Swe was the poorest performing subject, selecting the functional rake on only 6/10 trials. As this subject also took longer to reach criterion level at the two minute delay period, his performance may represent a cognitive limit in the time in which this subject can keep in mind the link between rake selection and reward retrieval events. Alternatively, it is possible this subject's motivation was decreased during experiments with longer delay periods, with less attention being given to the task resulting in fewer selections of the functional rake.

Table 5.11: Subjects' genus, age group and sex and number of correct selections of the functional rake across the 10 trials completed by subjects during Experiment 4 (significance on binomial test).

Subject	**Genus**	**Age group**	**Sex**	**Total number of correct trials**
Betty	*Hoolock*	Adult	F	9*
Chan Thar	*Hoolock*	Adult	F	10**
U Myint Swe	*Hoolock*	Sub-adult	M	6
Win Bo	*Hoolock*	Adult	M	10**
Iszie	*Hylobates*	Sub-adult	F	9*
Truman	*Hylobates*	Adult	M	9*
Violet	*Hylobates*	Sub-adult	F	10**
Asteriks	*Nomascus*	Adult	F	10**

Lucia	*Nomascus*	Sub-adult	F	9*
Ricky	*Nomascus*	Adult	F	8

* Significant at $p < 0.05$ level

** Significant at $p < 0.01$ level

5.3.7.1 Order effects

Mann-Whitney U tests were used to assess order effects for those enclosures where two individuals participated in testing, with time to first productive rake contact over the ten trials compared. There was no significant difference found between Win Bo and Chan Thar ($U = 42$, $z = -0.61$, $p = 0.58$), U Myint Swe and Betty ($U = 40.5$, $z = -0.72$, $p = 0.48$), or Asteriks and Lucia ($U = 31$, $z = -1.44$, $p = 0.17$). A significant difference was found between Violet and Truman in enclosure 9 ($U = 13$, $z = -2.8$, $p < 0.01$). The second individual to be tested, Truman, was faster to first productive contact (mean = 2.13 seconds) than Violet (mean = 4.29 seconds) but selected the functional rake less frequently (9/10 trials, Violet 10/10 trials). Testing order therefore did not appear to influence overall performance.

5.3.8 Comparison across delay periods in Experiments 1-4

Generalized Estimating Equations (GEE) were used to build a model assessing the effect of delay period, genus, age and sex on the number of correct selections of the functional rake (see Table 5.12). The model specifications were as described for assessment of performance levels during training (see section 5.2.4.1) with the following exceptions. Delay period, genus, age group and sex were set as factors. For parameter estimates, the five minute delay period was set as the reference category when evaluating performance differences across delay periods, *Hylobates* was the reference category when evaluating genera differences, with sub-adult the reference category when evaluating age differences, and female the reference category when evaluating sex differences. As the single *Symphalangus* subject who participated in training did not complete Experiments 1-4 she was excluded from the model. This left similar numbers of subjects from other genera who completed all experiments and so ‘genus’ was now included as a factor. As only ten trials were conducted at the five minute delay period, only the first ten trials from each delay period were assessed in order to give a comparison of performance over all delay periods for the first block of testing.

Table 5.12: GEE parameter estimates for factors used in model to compare frequency of functional rake selections between delay periods, genera, age groups and sexes.

Factor	**Parameter**	**B**	**SE**	**Lower CI**	**Higher CI**	**Wald Chi (χ^2)**	**df**	**Sig.**
Delay period	0 seconds	-0.06	0.06	-0.18	0.07	0.81	1	0.37
	30 seconds	-0.17	0.03	-0.06	0.06	0.00	1	1.00
	2 minutes	0.01	0.05	-0.08	0.11	0.05	1	0.82
	5 minutes	0[a]						
Genus	*Hoolock*	-0.03	0.04	-0.11	0.06	0.39	1	0.53
	Nomascus	-0.02	0.06	-0.15	0.10	0.13	1	0.72
	Hylobates	0[a]						
Age group	Adult	0.05	0.05	-0.04	0.14	1.35	1	0.25
	Sub-adult	0[a]						
Sex	Male	-0.04	0.05	-0.13	0.05	0.87	1	0.35
	Female	0[a]						

[a] Set to zero because this parameter is redundant

Delay period did not significantly predict performance (see Table 5.12); with similar levels of functional rake selections across experiments. Subjects selected the functional rake marginally more frequently during the two minutes delay experiment, with equivalent levels of functional rake selections across the 30 seconds and five minutes delay periods. Performance was generally poorer during the 0 second delay experiment, suggesting subjects may have been required to re-learn the association between rake selection and reward retrieval event once these were presented as temporally and spatially distinct. However, once subjects had learnt this association, the increasing time delay did not appear to significantly affect performance levels. Some subjects' performance began to decrease as the time delay increased (see Tables 5.9 and 5.10), suggesting the time periods may represent the cognitive limit within which these individuals could relate rake selection to reward retrieval. Alternatively, poorer performance may represent a loss of motivation or engagement with the task when longer delay periods were imposed. Further assessment with increased time delays is needed in order to assess the maximum time limit within which gibbons can successfully relate these two events.

Genus was not a significant predictor of functional rake selections (see Table 5.12); *Hylobates* subjects selected the functional rake slightly more frequently, with *Nomascus* and *Hoolock* subjects selecting the functional rake at equal levels. Genera differences in performance across previous chapters have varied. However *Hoolock* subjects have generally displayed higher levels of performance, selecting the functional rake more frequently than other genera, as has been found with previous assessments of object manipulation and tool use in gibbons (Cunningham, 2006). In contrast, *Hylobates* subjects showed a slightly improved performance during the current experiments. *Nomascus* and *Hoolock* species are found in more northern locations than *Hylobates* species, resulting in greater environmental variability (Brandon-Jones et al., 2004; Fan, Ai, Fei, Zhang & Yuan, 2013). Seasonal variation in both food availability and weather conditions may result in increased responsiveness to resources within *Nomascus* and *Hoolock* subjects. This increased responsiveness may have aided these subjects in previous experiments (see Chapters 3 and 4), where the reward was more immediately available, but resulted in a decreased performance once a delay was imposed in the current experiments, potentially as a result of less patient behaviour by these subjects. Neither sex nor age were significant predictors of performance (see Table 5.12); however, adults and females tended to select the functional rake more frequently than sub-adults or males, in line with findings from Chapters 3 and 4.

GEEs were used to build a further two models assessing the effect of delay period, genus, age and sex on the time taken to first enter the target area, and the time to first productive rake contact (see Table 5.13). Both models were as specified above for the model assessing performance levels, with the following exceptions. Both models were now specified as linear with identity link function, with either time to first approach or time to first productive contact as the dependent variable.

Table 5.13: GEE parameter estimates for factors used in models to compare the time to first enter the target area and time to first productive rake contact between delay periods, genera, age groups and sexes.

Model	**Factor**	**Parameter**	**β**	**SE**	**Lower CI**	**Higher CI**	**Wald Chi (χ^2)**	**df**	**Sig.**
Time to first enter target area	Delay period	0 seconds	-54.31	20.07	-93.66	-14.97	7.32	1	**<0.01**
		30 seconds	-14.52	23.56	-60.70	31.66	0.38	1	0.54
		2 minutes	-8.48	25.57	-58.60	41.64	0.11	1	0.74
		5 minutes	0[a]						
	Genus	*Hoolock*	-111.68	10.03	-131.34	-92.01	123.87	1	**<0.01**
		Nomascus	-88.07	12.04	-111.65	-64.48	53.54	1	**<0.01**
		Hylobates	0[a]						
	Age group	Adult	9.69	8.94	-7.83	27.21	1.17	1	0.28
		Sub-adult	0[a]						
	Sex	Male	5.66	9.91	-13.77	25.08	0.33	1	0.57
		Female	0[a]						
Time to first productive rake contact	Delay period	0 seconds	0.78	1.86	-2.87	4.42	0.17	1	0.68
		30 seconds	-0.15	0.72	-1.56	1.26	0.04	1	0.84
		2 minutes	0.13	1.26	-2.35	2.61	0.01	1	0.92

		5 minutes	0[a]						
	Genus	*Hoolock*	-0.33	1.08	-2.44	1.79	0.09	1	0.76
		Nomascus	-0.67	1.82	-4.23	2.89	0.14	1	0.71
		Hylobates	0[a]						
	Age group	Adult	1.23	1.20	-1.12	3.58	1.06	1	0.30
		Sub-adult	0[a]						
	Sex	Male	-2.25	0.94	-4.10	-0.40	5.70	1	**<0.05**
		Female	0[a]						

Significant results are in bold

[a] Set to zero because this parameter is redundant

Delay period significantly predicted time taken to first enter the target area surrounding platform 1 prior to rake selection (see Figure 5.4 and Table 5.13). The time taken generally increased as the delay period became longer, suggesting decreasing interest in the task. However, only the difference between Experiment 1 and Experiment 4 was significant ($p < 0.01$), with subjects being slower to enter the target area at the longest delay. All subjects reliably participated in testing, even at the five minute delay period, contrasting with findings of Harlow, Uehling and Maslow (1932), who found a gibbon (*H. lar*) was reluctant to respond to delays longer than a minute during a delayed response test. Delay period was not a significant predictor of time to first productive contact during rake selections (see Table 5.13). The time to first productive contact was similar across delay periods, with the increasing delay not resulting in longer durations spent considering the task before selection. This suggests subjects were not re-learning the association between rake selection and reward retrieval, but were holding in mind the relationship between these events, even when the temporal separation between them was five minutes.

Genus significantly predicted time to first approach platform 1 for rake selection with large variability between genera (see Figure 5.3), but not time to first productive contact at this platform (see Table 5.13). The difference between *Hoolock* and *Hylobates* ($p < 0.01$), and *Nomascus* and *Hylobates* ($p < 0.01$) was significant for time to first approach only. *Hylobates* were the slowest to enter the target area surrounding platform 1 and the slowest to first productive contact during rake selection. This slower, more considered, behaviour may have led to the slightly increased frequency of functional rake selections shown by this genus. *Hoolock* were fastest to first enter the target area, and *Nomascus* subjects were fastest to make productive contact. The longer latencies displayed by *Hylobates* subjects support Cunningham's (2006) findings that *Hylobates* subjects were slower to first solution during a raking in task. However, in contrast to Cunningham's (2006) findings, this slower behaviour resulted in slightly increased performance during the current task. As noted previously, differences in ecology between the tested genera may have influenced species differences in approach to the task, with the more stable environment of *Hylobates* species resulting in slower responses to potential resources (Fan et al., 2013).

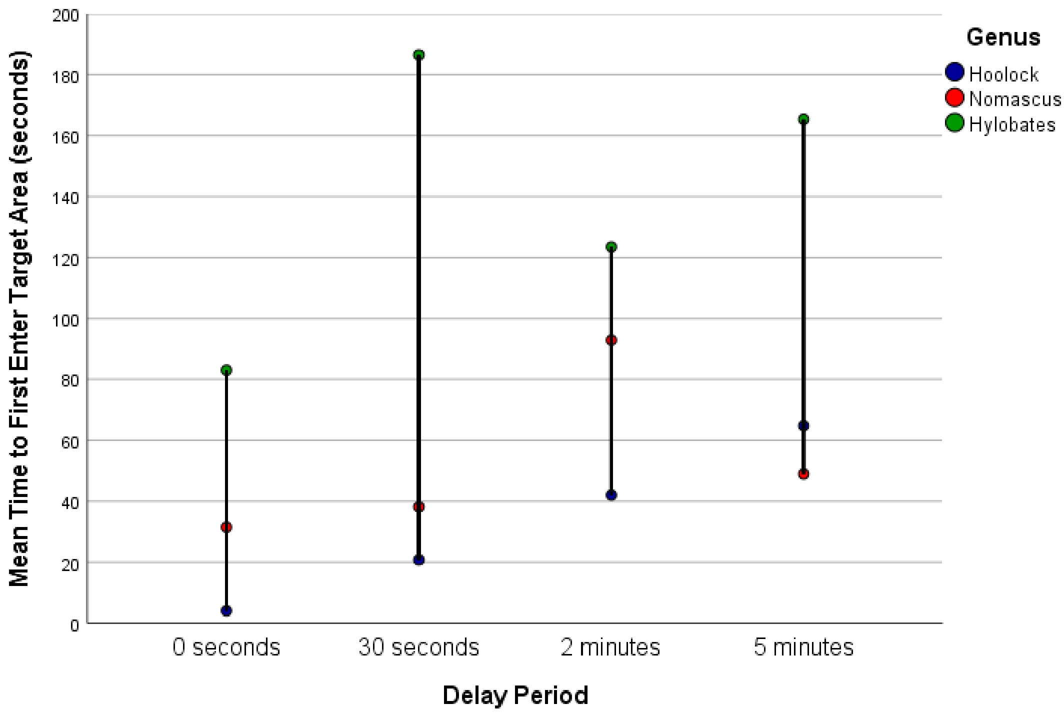


Figure 5.3: Mean time to first enter the target area surrounding platform 1 prior to rake selection for each genus during the first ten trials across delay periods.

Age was not a significant predictor across either model (see Table 5.13), with adults being only marginally slower to both enter the target area and to first productive contact during rake selections. Sex was a significant predictor of time to first productive rake contact only. Females were slower to first contact ($p < 0.05$) (see Table 5.13), but slightly faster than males to first enter the target area. These findings are consistent with the general trends shown across previous experiments, and support research (e.g., Cunningham, Anderson & Mootnick, 2011; Reader & Laland, 2001) suggesting that whilst males may be less risk-aversive as they have less to lose in fitness terms by investigating novel objects, females may display an increased attentiveness to prospective food sources given their increased energetic requirements (Cunningham et al., 2011).

A final model was built using GEEs to assess the effect of length of delay, genus, age, and sex on time spent attending to the task during the delay period (Experiments 2-4 only) (see Table 5.14). The model specifications were as described above for assessment of time to first approach and first productive contact, however, time spent attending to the task during the delay period was now set as the dependent variable. In order to control for overall length of delay which would affect opportunity for interacting with the apparatus (longer delays mean more time available to interact), time spent attending to apparatus during the delay period was calculated as a rate per minute for each experiment; thus, time spent attending during the 30 second delay period was multiplied by two, time spent at the two minute delay period was divided by two, and time spent at the five minute delay period was divided by five.

Table 5.14: GEE parameter estimates for factors used in model to compare time spent attending to the apparatus during the delay period between experiments (delay periods), genera, age groups and sexes.

Factor	Parameter	β	SE	Lower CI	Higher CI	Wald Chi (χ^2)	df	Sig.
Delay period	30 seconds	16.35	1.58	13.25	19.44	107.15	1	**<0.01**
	2 minutes	6.53	0.84	4.88	8.18	60.37	1	**<0.01**
	5 minutes	0[a]						
Genus	*Hoolock*	1.08	1.14	-1.15	3.31	0.91	1	0.34
	Nomascus	2.7	1.25	0.25	5.15	4.67	1	**<0.05**
	Hylobates	0[a]						
Age group	Adult	2.54	0.93	0.71	4.37	7.42	1	**<0.01**
	Sub-adult	0[a]						
Sex	Male	0.60	1.18	-1.70	2.91	0.26	1	0.61
	Female	0[a]						

Significant results are in bold

[a] Set to zero because this parameter is redundant

Delay length significantly predicted how long subjects spent attending to the apparatus during the delay period, with significant differences between the 30 second and five minute delay period ($p < 0.01$), and between the two minute and five minute delay periods ($p < 0.01$) (see Table 5.14). The average time spent attending to the task during the delay period decreased as the delay became longer (see Figure 5.4). As time to first enter the target area also became longer as the delay period increased, taken together, these results may suggest decreasing interest in the task. However, given that performance levels did not decrease during the trials with longer delays, these findings suggest gibbons were not required to attend to the apparatus longer during increased delays in order to keep in mind the association between the rakes and rewards.

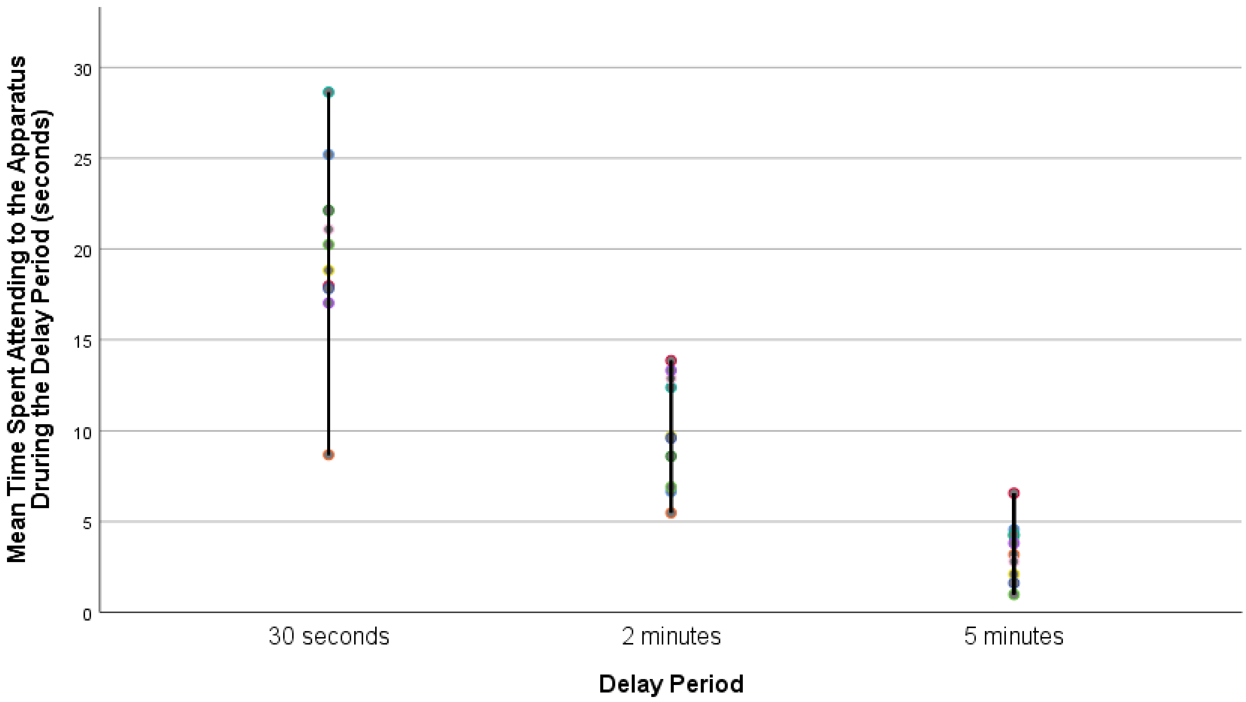


Figure 5.4: Time spent attending to the apparatus per subject during the first ten trials of each delay period, with times given as an average over 1 minute. Each coloured marker represents one subject.

There was a significant difference between *Nomascus* and *Hylobates* subjects ($p < 0.05$) in how long they spent attending to the task in the delay period (see Table 5.14). *Hoolock* (mean = 10.74 seconds) and *Nomascus* (mean = 12.36) subjects generally spent longer attending to the task than *Hylobates* subjects (mean = 9.66 seconds). This supports previous findings, suggesting *Hoolock* and *Nomascus* subjects were more responsive, potentially due to differences in ecology between these genera and *Hylobates*.

Age was a significant predictor, with adults spending longer attending to the task during the delay period ($p <0.01$) (see Table 5.14). This is consistent with the slightly longer latency to first productive contact displayed by adult subjects, suggesting these individuals were more engaged with the task. It is possible the absence of a food reward during the selection and delay periods of the current task contributed to the decreased interest in the task displayed by sub-adults. These subjects may not have kept in mind the association between the rake and future reward to the same degree as adult subjects, leading to decreased attentiveness to the apparatus during delay periods. In line with the findings that females were slower to first productive rake contact, female subjects also tended to spend less time attending to the task during the delay period, although this result was not significant (see Table 5.14).

5.4 Discussion

Gibbons successfully selected an un-baited functional rake needed to retrieve a food reward at a later time, doing so when the reward was presented immediately following rake selection during training, and when a time delay was introduced before reward retrieval at a second platform. Overall, the increase in delay period did not have a negative effect on subjects' ability to keep in mind the spatially and temporally separate events of rake selection and reward retrieval. Subjects required the greatest number of blocks to reach criterion level during training, with some evidence of subjects re-learning the association between the rakes and reward once the reward was not visually present during the selection period. Likewise, across the delay periods, subjects performed slightly poorer during Experiment 1, with no delay between the selected rake being taken to the retrieval platform, suggesting the implementation of a distinct spatial separation between rake selection and reward retrieval required some subjects to re-learn this association. However, the majority of subjects reached criterion level within the minimum of two blocks during training and Experiment 1. These findings may suggest a discontinuity in abilities between monkeys and gibbons, with Dekleva et al. (2012) findings subjects (*M. fascicularis*) displayed a context specific preference for rakes during an initial training set-up, presented on the testing platform with food rewards. Monkeys required progressive step-wise training in order to connect the distinct rake selection and reward retrieval events. In contrast, despite extended previous experience with the current rakes presented baited with food rewards (see Chapters 3 and 4), gibbons did not require training to readily select un-baited functional rakes without the presence of the food reward.

Performance across the remaining delay periods was relatively consistent, suggesting that once subjects had learned this association, they were capable of maintaining it over delay periods of five minutes. Some subjects displayed poorer performance as the delay period increased, potentially representing a cognitive limit within which these individuals could keep in mind the relationship between rake selection and reward retrieval. Further research is needed at longer delay periods in order to assess the limit within which gibbons are capable of associating temporally and spatially distinct events. Although performance across delay periods was consistent, the increasing wait period appears to have affected subjects' motivation levels, as they generally took longer to enter the target area surrounding the selection platform as the delay increased. Despite this increased latency to approach the task, there was no significant difference in time to first productive rake contact across the different delay periods. Subjects therefore did not appear to require longer considering the task before making selections as the delay period increased. The time subjects spent attending to the task during the delay period decreased as the delay increased; suggesting subjects were also not required to return to the apparatus more frequently in order to keep in mind the association between the rakes and rewards at longer delays.

Hoolock subjects have displayed a trend for increased frequency of functional rake selections across previous chapters. Due to the small number of subjects, interpretations of the current findings are cautious; however, general trends in performance differences seen in the current experiments may be linked to differences in species ecology. *Hylobates* species have lower levels of seasonality in diet, being found in more southern locations than *Hoolock* or *Nomascus* species (Brandon-Jones et al., 2004), with food availability being more stable. This may contribute to the less impulsive, or impatient, behaviour displayed by *Hylobates* subjects here and in previous research (see Cunningham, 2006) compared to more northern species that may have to be more opportunistic and investigative when faced with novel problems in resource acquisition. Here, *Hylobates* were significantly slower to first approach the task than the other genera during Experiments 1-4. These subjects also spent the least time attending to the task during the delay period, suggesting less impatience with the task. The slower behaviours displayed by *Hylobates* subjects contrast with previous chapters, with *Nomascus* subjects generally spending the longest time attending to the task and in reward directed rake contact, and being slower to first productive rake contract across the experiments in Chapter 4. This is consistent with the suggestion that increased seasonal variability experienced by *Nomascus* subjects may result in an increased responsiveness to potential food sources, with the absence of the food reward during rake selection potentially resulting in decreased interest in the task by *Nomascus* subjects.

Previous research has found either no difference in performance levels between age groups in gibbons (Cunningham, 2006; Fedor, Skollár, Szerencsy & Ujhelyi, 2008), or an increased performance by sub-adults (Cheyne, 2009). This contrasts with suggestions that the increased experience of adult primates allows for more efficient behaviour and increased problem solving abilities (Griffin & Guez, 2014; Kendal et al., 2005; Reader & Laland, 2001). There were no significant differences in performance between age groups in the current experiments; however, the general trend for adult subjects outperforming sub-adults across Chapter 4 was also apparent during Experiments 1-4. Adults were significantly slower than sub-adults to first productive rake contact during training, with this trend also found across Experiments 1-4. These findings are in line with those of previous chapters, but contrast with those of Kendal et al. (2005) who found adult *Callitrichid* monkeys displayed shorter latencies to first successful manipulations during an extractive foraging task than younger individuals. The finding that adults spent significantly longer attending to the task during the delay periods also contrasts with Kendal et al.'s (2005) finding that attentiveness to a foraging task decreased as age increased. It is possible that differences in the type of task resulted in the decreased attentiveness displayed by sub-adults in the current experiments. The current task required subjects to keep in mind the relationship between the rakes and rewards over the time delay, with the food reward not visible until the retrieval period, in contrast to the foraging task. It is possible sub-adults were less successful at keeping the future reward in mind than adults, with this resulting in a lack of interest in the un-baited apparatus.

Sex differences were similar to previous chapters, with females tending to select the functional rake at higher frequencies than males, although not significantly so. Females were significantly slower than males to first productive rake contact during Experiments 1-4. This longer latency to interact with the apparatus supports previous research in suggesting females may be more risk-aversive, with these individuals having more to lose, in fitness terms, by investigating novel objects (Cunningham et al., 2011; Reader & Laland, 2001). All subjects were successful during Experiments 1-4; however, during Chapter 4, where a number of subjects failed to reach criterion level, successful subjects tended to display slower behaviours than unsuccessful subjects. The increased latency to first contact by adults and females may suggest more time spent considering the task by these subjects, potentially aiding in the slightly improved performance displayed by these two groups.

In line with Suddendorf and Corballis's (2007; 2010) criteria for assessment of prospective cognition, the current task presented a temporal and spatial separation between action (rake selection) and response (reward retrieval), aiming to limit behavioural cueing. However, subjects had likely formed associations between the functional rake and food reward throughout the multiple trials presented using the same apparatus in previous testing (see Chapters 3 and 4). The retrieval table was also visible during rake selection, potentially providing a more subtle behavioural cue. Thus, subjects' behaviour may have been guided by these learned associations between the present stimuli and future reward, rather than mentally representing the future reward event. Current findings therefore display evidence of basic prospective cognition within gibbons, displaying the ability to maintain an association between rake selection and reward retrieval when this event was delayed up to periods of five minutes. Learned associations such as this are prospective in the sense that subjects are guiding future actions by learning which behaviours lead to this action (Suddendorf et al., 2018).

As has been found within large apes, gibbons therefore displayed the ability to select functional tools for future use. However, in contrast to the findings of Osvath and Osvath (2008), and Mulcahy and Call (2006), the current task is not considered as requiring a sense of the future (Raby & Clayton, 2009). Large apes have displayed the ability to transport functional tools over extended time delays and to attend to the particularities of future reward retrieval events, selecting novel tools which are functional for future tasks (Osvath & Osvath, 2008). Future research minimising the applicability of associative learning accounts for subjects' behaviour is needed in order to assess whether gibbons are capable of higher levels of prospective cognition displayed by large apes; however, the aim of the current research was to assess the presence of basic prospection in gibbons. As noted by Raby and Clayton (2009), investigations of prospective cognition not requiring a sense of the future have been overlooked in favour of assessing the human uniqueness to mentally travel in time. A broader focus, including assessment of component processes of prospective cognition allows for a clearer view of the emergence of the capabilities across species. In line with this proposal, Chapter 6 presents an assessment of self-control capacities within gibbons, an important component of prospective cognition without which individuals would be unable to inhibit current desires in favour of future ones.

Chapter 6: Self-control for a future reward

6.1 Introduction

Chapter 5 presented an assessment of basic prospective cognition in gibbons (*Hylobatidae*), with subjects successfully selecting functional rakes for future use. In line with Raby and Clayton's (2009) suggestion that broader investigations of prospective cognition and component processes are needed in order to present a clearer overview of the emergence of capabilities in primates, Chapter 6 presents an initial investigation of self-control capacities within gibbons. Self-control is a crucial component of prospective cognition, needed in order to plan for future desires outside the scope of current motivations. Regardless of whether an individual is capable of displaying levels of prospective cognition requiring a sense of the future or not (as in episodic prospection), actions towards future outcomes cannot be taken without first overcoming conflicting psychological states (Osvath & Martin-Ordas, 2014). Tasks requiring subjects to display self-control, foregoing an immediate reward in order to obtain a more valuable delayed reward, set current and future desires in opposition by design. A variety of tasks have been used to assess self-control in nonhuman primates. The use of inter-temporal choice tasks assesses a subjects' self-control by choosing between a more immediately available small reward or a large reward which is only accessible following a delay period. However, as both rewards are routinely visible during these tasks, and animals have displayed difficulties with inhibiting selecting a larger quantity of food when presented with small and large food options simultaneously, it is not clear whether subjects' selections of the large reward represent a difficulty in inhibiting this behaviour or intention to wait for the larger reward (Beran et al., 2014, see Shifferman, 2009, for a review of reverse reward contingency research).

Accumulation tasks may circumvent this issue by assessing delay maintenance; the ability to continually forego an immediate reward, instead waiting to retrieve a larger reward at a later time (Beran, 2002). These tasks do not present the choice between two reward quantities, but assess subjects' ability to delay gratification until a maximum reward quantity has accumulated. A supply of food rewards are placed one at a time within subjects reach, with the addition of food rewards ending if the subject chooses to eat any of the pieces. Beran and Evans (2006) found chimpanzees (*Pan troglodytes*, n =4) tolerated delays of up to 11 minutes during an accumulation task. Chimpanzees also waited for a larger reward when they could not see how many more food items could be transferred by an experimenter, and did so when an automated system transferred rewards, in the absence of a human experimenter.

In contrast, Evans and Beran (2007) found only three out of nine tested rhesus macaques (*Macaca mulatta*) consistently exhibited delay maintenance greater than 15 seconds during an initial accumulation task. One piece of a preferred food was then included in the transferred items, being positioned either at the start, middle, or end of the order of transferred items. Monkeys were found to delay gratification longer when the preferred food item was transferred later, with Evans and Beran (2007) suggesting subjects' self-control may be inversely related to reward quality. A further experiment presented subjects with a choice between a single immediately available but less preferred reward, and a single, delayed preferred reward. Four of the nine subjects displayed delay maintenance on at least one trial with a delay period of two minutes; however, the majority of subjects failed to wait for the preferred reward. Anderson, Kuroshima and Fujita (2010) report similar findings assessing both capuchin (*Cebus apella*, n = 4) and squirrel monkeys (*Saimiri sciureus*, n = 4) during an accumulation task with a maximum delay of 30 seconds. Only one squirrel monkey displayed any ability to delay gratification during an initial assessment. However, further testing found an additional squirrel monkey and two capuchin monkeys capable of delaying gratification once the food rewards increased in size as the transfer continued, with Anderson et al. (2010) suggesting increasing size may have acted as a facilitator, assisting subjects in understanding the task demands. The levels of self-control displayed by monkeys are therefore lower than those found in chimpanzees, with monkeys only tolerating short time delay periods during accumulation tasks.

Self-control tasks incorporating tool use also avoid the difficulties of selecting a smaller reward inherent in inter-temporal choice tasks. By removing the presence of the small and large quantities of food rewards these tasks remove the potential for the larger reward stimulating a response to select this quantity. Evans and Westergaard (2006) presented 20 capuchin monkeys (*C. apella*) with edible rod shaped tools, which could be used as a dipping stick to retrieve a preferred reward. In order to maximise reward retrieval, subjects would be expected to retain and transport the tool, retrieve the preferred reward, and then consume the tool. The monkeys were categorised based on previous experimental experience. Five subjects were categorised as highly experienced, having participated in two or more previous tool using experiments. A further five subjects had participated in one previous investigation of tool use and were categorised as moderately experienced, and the remaining ten subjects had no previous experience. The value of the tool was varied, being either low value celery or high value pretzel, and the preferred reward, a cylinder containing peanut butter, was presented either 1 metre or 5 metres from the location where subjects were handed tools. Neither tool quality or reward distance were found to significantly affect subjects ability to display self-control, with Evans and Westergaard (2006) suggesting the difference in quality and location may not have been sufficient enough to produce differences in behaviour.

The level of self-control displayed related to subjects' previous experience, with the highly experienced subjects retrieving the preferred reward more frequently than moderately experience subjects or those with no experience. Highly experienced subjects retrieved the preferred reward on all trials except those where a high value tool was presented with the reward located at a distance of 5 metres, suggesting the greater energy expenditure required to retrieve the preferred reward was outweighed by the immediately consumable tool that was of high value. The moderately experienced subjects did not discriminate between tool quality, but only retrieved the preferred reward when it was positioned at a distance of 1 metre, suggesting for these subjects, the consumable tool was more desirable than a reward which required more effort to attain. Those subjects with no previous experience performed poorest, using the tool to retrieve the reward only when presented with a low value tool and a reward positioned at a distance of 1 metre. It is possible the inexperienced subjects did not fully understand the behavioural requirements of the dipping task. However, Evans and Westergaard's (2006) proposal that tool using abilities relate to levels of self-control is supported by the finding that, in contrast to highly experienced subjects, moderately experienced subjects only retrieved the preferred reward when this was positioned at the closer distance of 1 metre.

Evans and Westergaard (2006) further propose the relationship between tool use and self-control may be due to both behaviours requiring foresight and deliberate control over ones actions. Individuals with more tool using experience may develop higher levels of self-control through foregoing immediate rewards in order to use or manufacture tools to obtain preferred food sources. Support for this idea comes from Osvath and Osvath (2008) who assessed self-control in relation to foresight during a series of tool using tasks with chimpanzees (*P. troglodytes*, n = 2) and one orangutan (*Pongo abelii*) (see Chapter 2, section 2.2.2 for a more detailed discussion). Subjects were presented with a choice between a functional tool, usable to retrieve a larger future reward, three distractor objects and a small food reward. The apes displayed the ability to override current desires in favour of satisfying future needs, refraining from selecting the immediate small reward in order to select the functional tool for future reward retrieval on the majority of trials.

Osvath and Osvath (2008) suggest the future desire, placed in conflict with current desires during self-control tasks, may be related to a pre-experience of the future situation. However, they outline certain experimental design requirements that must be met in order to assess the ability to plan for future needs. Firstly, along with presenting rewards of different quantity, rewards differing in quality should be presented to produce competition between different current and future desires. For example, during the selection period of Osvath and Osvath's (2008) assessment of self-control (see above), an immediate reward of a piece of fruit was offered along with a functional tool usable at a later time to retrieve a half a litre of fruit juice. Thus, when presented with this choice, subjects capable of planning for future needs may represent the future need for the fruit juice, and this desire may be placed in opposition with the current desire for the piece of fruit. Secondly, the stimuli related to the future event should not have an associatively learned connection with the future reward; that is, the future oriented behaviour should not be cued by the presence of a stimulus with an associative relation to the future event. Although cognition resulting in prospective behaviour may be cued, either internally or externally, associative learning is characterised by arbitrary connections between stimulus and reward and may not require mentally representing future situations.

The current research aimed to assess if gibbons were capable of foregoing an immediate reward, instead selecting a functional rake in order to retrieve a larger reward. Although gibbons are not prolific tool users, with few reports of tool use in wild gibbons (see Shumaker, Walkup & Beck, 2011), the current subjects had completed various object manipulation experiments (see Chapters 3, 4 and 5). These experiments were not classified as true tool use, in that subjects were not required to properly orient the tool before raking in a reward (Shumaker et al., 2011). However, in line with Evans and Westergaard's (2006) proposal that tool use abilities may relate to self-control, it is possible object manipulation experience may aid subjects in displaying inhibitory control. Given gibbons phylogenetic positon, and that some monkeys with prior tool using experience successfully displayed self-control in order to retrieve a preferred reward (Evans & Westergaard, 2006), it was expected subjects would inhibit their desire for the immediate reward, selecting the functional rake to retrieve a later larger reward. In line with Osvath and Osvath's (2008) criteria for assessing self-control in planning paradigms, qualitatively different rewards were presented, inducing different current and future desires. However, given subjects previous experience with this task (see Chapters 3, 4 and 5), and the number of trials given throughout, it is possible subjects had learned associations between the functional rake and reward retrieval. If there is an associatively learned connection between the tool and the future reward, the tool would carry an intrinsic value detached from cognitive considerations about possible future rewards. An additional experiment therefore assessed whether subjects' selections were guided by an association with the rake, or whether their behaviour was reward directed.

6.2 Training

6.2.1 Subjects

Subjects included eight gibbons housed at the Gibbon Conservation Center (see section 3.2.1) that had selected the functional rake significantly above chance levels across the experiments of Chapter 5. This included three eastern hoolock (*Hoolock leuconedys*), two northern white cheeked (*Nomascus leucogenys*), and two pileated (*Hylobates pileatus*) gibbons (see Table 6.1). Subjects were separated and tested individually. As subjects in enclosures 6, 9 and 15 may have witnessed other individuals participating in testing, order effects were calculated for these individuals during training trials, with the order of testing not found to affect performance levels (see section 6.2.3).

Table 6.1: Subjects' species, age, sex, housing information. Number of times each subject selected the large reward during the 10 'reward size preference' trials, and whether they reached criterion level performance during training.

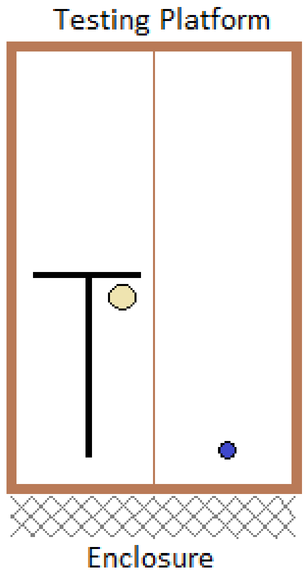
Subject	Enclosure	Species	Age	Sex	Group	Number of times large reward selected during reward preference trials	Reached criterion during training
Chan Thar	6	*H. leuconedys*	10	F	Male/female pair	9	✓
Win Bo	6	*H. leuconedys*	13	M	Male/female pair	10	X
Betty	10	*H. leuconedys*	17	F	Male/female pair	10	✓
Iszie	5	*H. pileatus*	4	F	Family	8	X
Truman	9	*H. pileatus*	13	M	Male/female pair	10	✓
Violet	9	*H. pileatus*	7	F	Male/female pair	10	✓
Asteriks	15	*N. leucogenys*	16	F	Family	8	✓
Lucia	15	*N. leucogenys*	6	F	Family	8	✓

6.2.2 Experimental procedure and apparatus

In order to confirm subjects' preference for a larger reward, they were first presented with ten trials offering a choice between a small (one blueberry) and large (piece of sweet potato) reward (see section 6.2.3). The large reward was substantially bigger than the blueberry, being approximately the size of 5-10 blueberries. The rewards were placed on opposite sides of a testing platform located outside subjects' enclosures (as described in section 3.2.2), 60cm from the platform edge closest to the enclosure, and therefore out of subjects' reach (see Figure 6.1). The position of the large reward, on either the left or right side of the platform was randomised across trials, appearing on both sides equally but never on the same side for more than two consecutive trials during one block. Following baiting, a five second delay occurred before subjects could select one of the two rewards by reaching their hand towards the chosen reward. To be considered as selecting a reward, the subject needed to pass their hand over a line marked 20cm back from the testing platform edge on either the left or right of the platform corresponding to the position of the food item chosen. The reward on the selected side was then moved into reach by the experimenter. A 20 second inter-trial interval was used before the next trial began. A note of large reward position (left/right) and which reward was selected (small/large) were taken during testing.

Figure 6.1: Northern white cheeked gibbon (*N. leucogenys*) selecting the large reward during the reward size preference trials.

Once gibbons had completed the reward preference trials, the training stage aimed to assess whether subjects were able to inhibit a desire for an immediate small reward, instead selecting a rake usable to obtain a larger reward. A small reward, and functional rake (F1 as described in section 3.2.2) baited with a large reward were presented out of reach, on opposite sides of the testing platform positioned outside subjects' enclosures. A blueberry was again used as the small reward and a piece of sweet potato as the large reward. The large reward was placed on the testing platform, out of reach at a distance of 60cm from the platform edge. A five second delay then took place before the small reward and rake handle were pushed within reach, with both rake handle and small reward being 20cm back from the platform edge closest to the enclosure and therefore accessible to the gibbon. The large reward remained out of reach, retrievable only by pulling in the rake to slide the reward within reach (see Figure 6.2). Subjects were able to make a selection between the large or small reward by means of passing their hand over a line marked 20cm back from the platform edge, either retrieving the blueberry or pulling-in the rake. The non-selected reward was immediately removed from the platform following selection.



= Large reward. = Small reward. = Functional rake.

Figure 6.2: Experimental set up when gibbon made a choice, showing the functional rake, large and small rewards positioned on the testing platform.

Subjects were presented with a maximum of 120 trials, in blocks of ten trials per day with an inter-block interval of a minimum of five minutes being enforced. Criterion for passing to the next experiment was 8/10 selections of the large reward on two consecutive blocks. The position of the large reward on either the left or right side of the platform was randomised, appearing on both sides in equal numbers but never more than twice on the same side successively during a block. All testing was recorded using a video camera for later behavioural coding. A note of the large reward position (left/right) and reward selected (large/small) was taken during testing. The time to first contact, with either the rake or the small reward, was coded from the testing videos. Time to first contact was taken from the time when subjects entered the target area and attended to the apparatus until first contact was made, assessing subjects' impulsivity in making selections.

6.2.3 Results

The number of times subjects selected the larger piece of sweet potato during the reward size preference trials varied from 8/10 to 10/10 trials (see Table 6.1). All subjects were therefore considered as displaying a preference for the sweet potato over the smaller reward of one blueberry and progressed to the training stage. Six of the eight subjects reached criterion level performance in the testing phase, taking between two and seven blocks to do so (see Table 6.2). Performance across all trials completed by each subject was evaluated using binomials tests. Win Bo and Iszie did not select the large reward significantly above chance levels across all test trials; these two subjects failed to reach criterion level and were excluded from further testing. Chan Thar and Violet also failed to select the large reward at significant levels across all blocks; however, these two subjects reached criterion level performance in their final two blocks and so progressed to Experiment 1. Subjects who reached criterion level took longer to first contact either the rake or small reward (mean = 3.25 seconds) than those who were unsuccessful (mean = 2.56 seconds), suggesting slower, more considered behaviour resulted in better performance.

Table 6.2: Subjects' genus, age group and sex. Number of selections of the large reward in each block completed by subjects (1 block = 10 trials), across the total number of trials completed by each subject, and the two criterion blocks (significance on binomial test). Blocks where subjects reached criterion level performance are highlighted.

Subject	Genus	Age group	Sex	Total number of large reward selections per block												Total number of large reward selections	Performance on criterion blocks
				1	2	3	4	5	6	7	8	9	10	11	12		
Betty	*Hoolock*	Adult	F	9	10											19/20**	19/20**
Chan Thar	*Hoolock*	Adult	F	5	5	8	8									26/40	16/20*
Win Bo	*Hoolock*	Adult	M	1	3	5	5	0	0	0	0	0	2	3	3	22/120	n/a
Iszie	*Hylobates*	Sub-adult	F	3	2	5	6	6	8	6	4	5	5	6	7	63/120	n/a
Truman	*Hylobates*	Adult	M	6	5	7	10	6	10	8						52/70**	18/20**
Violet	*Hylobates*	Sub-adult	F	3	4	6	10	9								32/50	19/20**
Asteriks	*Nomascus*	Adult	F	2	6	7	8	10								33/50*	18/20**
Lucia	*Nomascus*	Sub-adult	F	7	10	9										26/30**	19/20**

■ = Criterion level performance.

* Significant at $p < 0.05$ level.

** Significant at $p < 0.01$ level.

6.2.3.1 Order effects

Order effects were assessed using Mann-Whitney U tests for those enclosures where two individuals participated in testing, with time to first contact over the first ten trials compared. There was no significant difference found between Win Bo and Chan Thar in enclosure 6 ($U = 49.5$, $z = -0.04$, $p = 0.97$), Truman and Violet in enclosure 9 ($U = 26$, $z = -1.82$, $p = 0.08$), or Asteriks and Lucia in enclosure 15 ($U = 46$, $z = -0.3$, $p = 0.8$), suggesting testing order did not affect performance.

6.3 Experiment 1

6.3.1 Subjects

The six subjects that reached criterion level performance during training participated in Experiment 1 (see Table 6.3). Subjects were separated and tested individually. Order effects were calculated for subjects in enclosures where two individuals were tested, as these subjects may have witnessed other individuals participating in testing (see section 6.3.3.1)

Table 6.3: Subjects' species, age, sex and housing information.

Subject	Enclosure	Species	Age	Sex	Group
Chan Thar	6	*H. leuconedys*	10	F	Male/female pair
Betty	10	*H. leuconedys*	17	F	Male/female pair
Asteriks	15	*N. leucogenys*	16	F	Family
Lucia	15	*N. leucogenys*	6	F	Family
Truman	9	*H. pileatus*	13	M	Male/female pair
Violet	9	*H. pileatus*	7	F	Male/female pair

6.3.2 Experimental procedure and apparatus

This experiment aimed to assess if subjects could continue to reject an immediately attainable small reward, instead selecting a rake usable to obtain a large reward located on a second testing platform, not in direct relation with the rake. The apparatus used during Experiment 1 consisted of two testing platforms (as described in section 3.2.2), identical other than the presence of a wooden dividing strip on the first acting as a middle divider, one small reward (blueberry), one large reward (piece of sweet potato), and a functional rake (F1, as described in section 3.2.2). The procedure was identical to that used in training (see section 6.2.2) with the following exceptions. The rake remained un-baited during the selection period, with the large reward being placed, out of reach (60cm from platform edge), on platform 2 before the rake and small reward were placed on platform 1 (see Figure 6.3). The rake handle and small reward were placed on platform 1 at a reachable distance of 20cm from the platforms edge closest to the gibbon immediately following placement of the large reward on platform 2. The distance between the two testing platforms varied due to differences in enclosure location and size; however, platform 2 (and the large reward) were always positioned within view of platform 1. A selection was counted when subjects passed their hand over a line marked 20cm back from the edge of platform 1, on either the side containing the rake or the small reward. The non-selected item was immediately removed from the platform following selection. Subjects could pick up the small reward if this was selected. If the rake was selected, it was immediately taken to platform 2 and placed, baited, within reach, allowing subjects to use the rake to retrieve the larger reward. Subjects were given a maximum of 60 trials (six blocks), with criterion level classified as 8/10 selections of the rake, and therefore large reward, on two consecutive blocks.

Testing was recorded with a video camera for later behavioural coding, with a note of large reward position (left/right) and reward selected (large/small) taken during testing. The time to first contact, with either the rake or the small reward was coded from the testing videos. This was taken from the time when subjects entered the target area and attended to the apparatus until first contact was made.

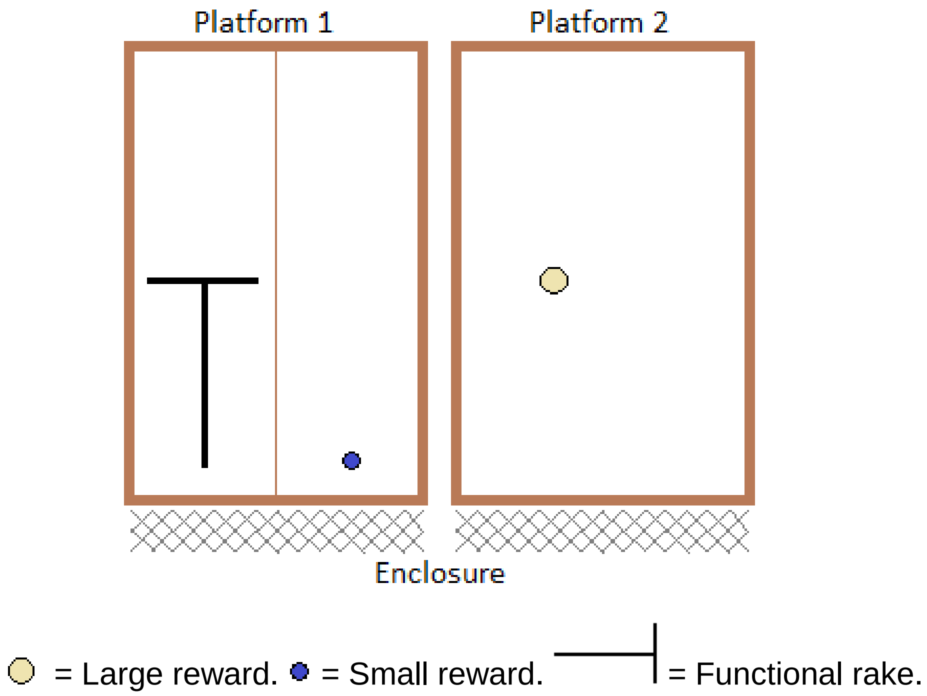


Figure 6.3: Experimental set up when gibbon made a choice, showing the functional rake and small reward positioned within reach at a distance of 20cm from the subject on testing platform 1, and the large reward placed out of reach at a distance of 60cm on platform 2.

6.3.3 Experiment 1 results

Subjects preferentially selected the small reward, with very few selections of the large reward and no subject reaching criterion level performance within the maximum of six blocks (see Table 6.4). The time taken to first contact with either the rake or small reward decreased across blocks, with subjects displaying more impulsive behaviour as testing progressed, potentially not considering both options before making a selection (see Figure 6.4).

Table 6.4: Subjects' genus, age group and sex. Number of selections of the large reward in each block completed by subjects (1 block = 10 trials), and across the total number of trials completed by each subject.

Subject	**Genus**	**Age group**	**Sex**	**Total number of large reward selections per block**						**Total number of large reward selections**
				1	**2**	**3**	**4**	**5**	**6**	
Betty	*Hoolock*	Adult	F	0	0	0	1	0	0	1/60
Chan Thar	*Hoolock*	Adult	F	0	1	0	0	0	0	1/60
Truman	*Hylobates*	Adult	M	5	2	3	2	4	6	22/60
Violet	*Hylobates*	Sub-adult	F	0	1	1	1	0	0	3/60
Asteriks	*Nomascus*	Adult	F	2	3	2	2	2	2	13/60
Lucia	*Nomascus*	Sub-adult	F	0	5	1	4	2	1	13/60

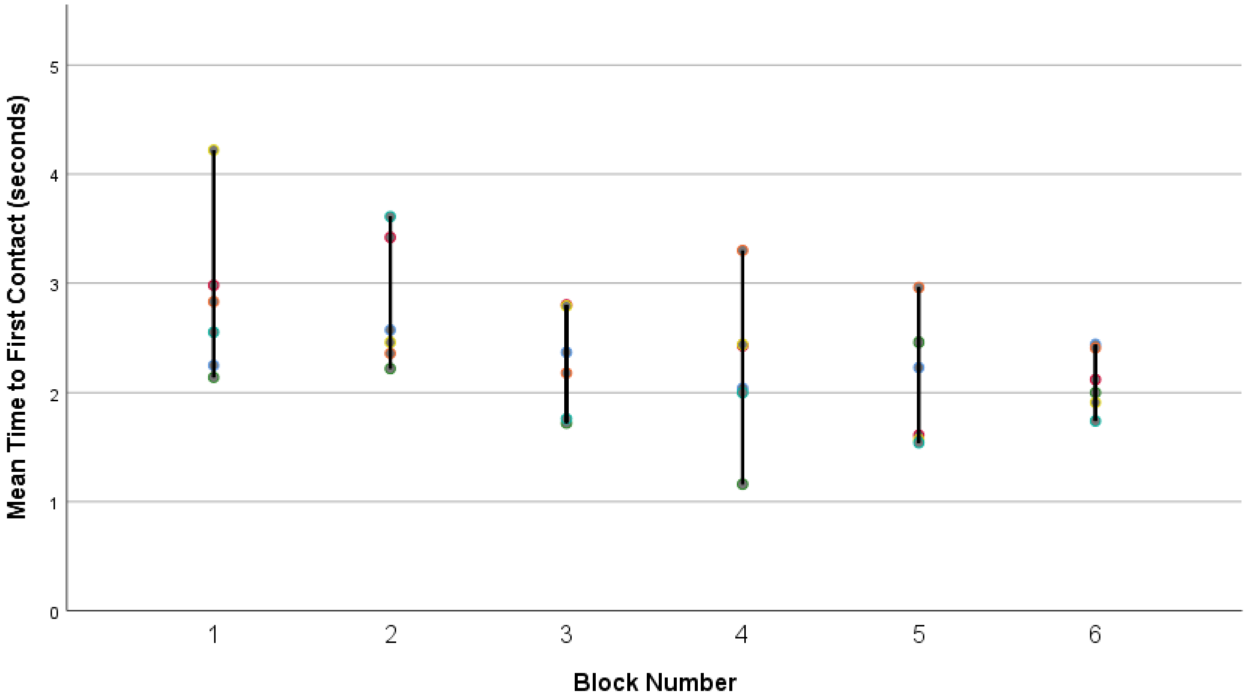


Figure 6.4: Mean time to first contact with either the rake or small reward for subjects across blocks 1 to 6. Each coloured marker represents one subject.

Findings suggest subjects lacked the ability to inhibit selection of the immediately attainable small reward, in contrast to findings of Osvath and Osvath (2008) and Evans and Westergaard (2006), who found subjects would forego more immediate rewards in order to obtain delayed larger rewards. It was expected, given subjects previous experience of selecting a rake at one platform in order to retrieve a reward at a second platform (see Chapter 5), that subjects would display an understanding of the rake being useable for large reward retrieval. However, it is possible the introduction of a small reward led to a lack of understanding of the task demands, potentially being unaware that the rake would be taken to second platform if selected. The following experiment therefore aimed to make this relationship clear to subjects.

6.3.3.1 Order effects

Mann-Whitney U tests were used to assess order effects for enclosures 9 and 15, where two individuals were tested, with time to first contact over the first ten trials compared. No significant difference was found between Violet and Truman ($U = 25$, $z = -1.89$, $p = 0.06$), or Lucia and Asteriks ($U = 45$, $z = -0.38$, $p = 0.74$), suggesting testing order did not have an effect on performance.

6.4 Reinforcement experiment

6.4.1 Experimental procedure, apparatus and subjects

Subject's poor performance during Experiment 1 may have been due to a lack of understanding that the rake presented on testing platform 1 would be taken to platform 2 if selected. The decreased time to first contact across testing in Experiment 1 suggests subjects were also not considering the two options before making a selection, possibly due to a lack of understanding of the task demands. This stage therefore aimed to reinforce that the rake would be moved to platform 2, being usable to retrieve the large reward. Subjects were required to choose between selecting the easier obtainable small reward, or selecting, and then using, the rake to pull in the large reward. The same six subjects that participated in Experiment 1 were tested (see Table 6.5), with all subjects again separated and tested individually.

Two testing platforms were again presented, with the rake and large reward both placed on platform 2; 60cm from the platform edge and therefore out of subjects reach. The small reward was then immediately placed on platform 1, also at a 60cm distance from the platform edge (see Figure 6.5). The testing platforms, rake and rewards were as described in section 6.3.2. Following baiting, subjects were able to make a selection by passing their hand over a line marked 20cm back from the either platform's edge. If the small reward was selected, it was immediately pushed to within reach by the experimenter. If subjects selected the large reward, the rake was pushed to within reach, with subjects then able to pull in the rake and retrieve the reward. Criterion was again set at 8/10 selections of the large reward over two consecutive blocks, with a maximum of 120 trials permitted. A note of the large reward position (left/right) and reward selected (large/small) was taken during testing.

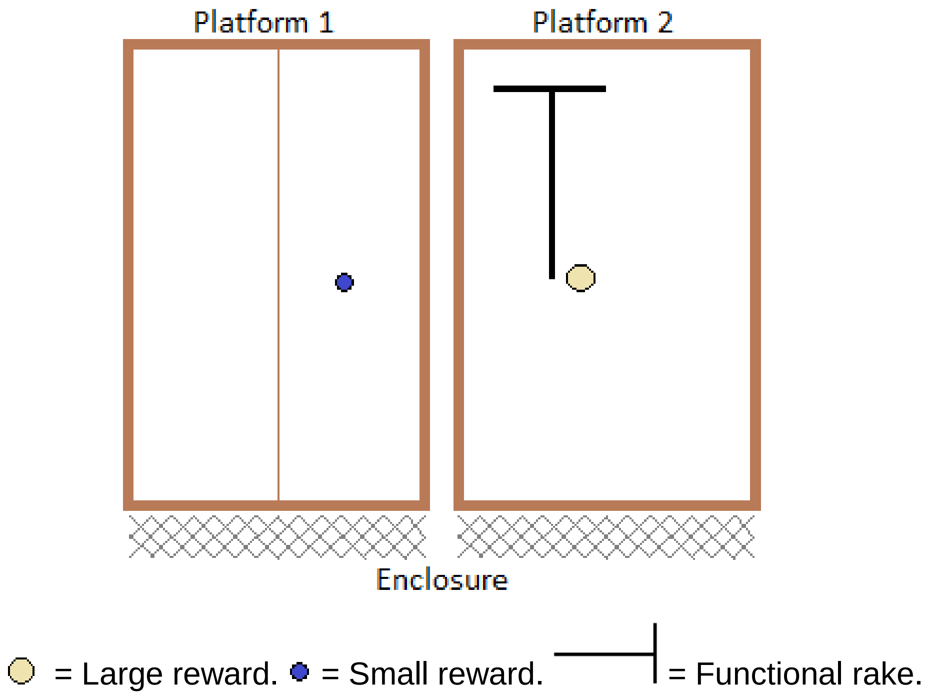


Figure 6.5: Experimental set up when gibbon made a choice, showing the small reward placed on platform 1, and the rake and large reward positioned on platform 2. All items are positioned out of reach at a distance of 60cm from the subject.

6.4.2 Reinforcement experiment results

All six subjects reached criterion level performance within two blocks, with all selecting the large reward significantly above chance levels (see Table 6.5).

Table 6.5: Subjects' genus, age group and sex. Number of selections of the large reward in each block completed by subjects (1 block = 10 trials), and across the total number of trials completed by each subject (significance on binomial test).

Subject	Genus	Age group	Sex	Total number of large reward selections per block		Total number of large reward selections
				1	2	
Betty	*Hoolock*	Adult	F	10	10	20/20**
Chan Thar	*Hoolock*	Adult	F	9	8	17/20**
Truman	*Hylobates*	Adult	M	10	10	20/20**
Violet	*Hylobates*	Sub-adult	F	10	8	18/20**
Asteriks	*Nomascus*	Adult	F	9	10	19/20**
Lucia	*Nomascus*	Sub-adult	F	9	10	19/20**

** Significant at $p < 0.01$ level.

6.5 Experiments 2-4

6.5.1 Subjects

The six subjects that participated in Experiment 1 and the reinforcement experiment participated in Experiment 2 (see Table 6.6). Following this, subjects were required to reach criterion level in order to progress to the next experiment. All six subjects reached criterion level in Experiment 2 and progressed to Experiment 3. Three subjects failed to reach criterion level during Experiment 3 and so were excluded from Experiment 4. These three subjects participated in further experiments detailed in section 6.9. Subjects were separated and tested individually. Order effects were calculated for subjects in enclosures where two individuals were tested, as these subjects may have witnessed other individuals participating in testing (see individual experiments results sections for analyses).

Table 6.6: Subjects' species, age, sex, housing information, and whether they reached criterion level performance or not across Experiments 2, 3, and 4.

Subject	Enclosure	Species	Age	Sex	Group	Exp. 2	Exp. 3	Exp. 4
Chan Thar	6	*H. leuconedys*	10	F	Male/female pair	✓	X	
Betty	10	*H. leuconedys*	17	F	Male/female pair	✓	X	
Asteriks	15	*N. leucogenys*	16	F	Family	✓	X	
Lucia	15	*N. leucogenys*	6	F	Family	✓	✓	✓
Truman	9	*H. pileatus*	13	M	Male/female pair	✓	✓	✓
Violet	9	*H. pileatus*	7	F	Male/female pair	✓	✓	✓

✓ = Reached criterion level.

X = Did not reach criterion level.

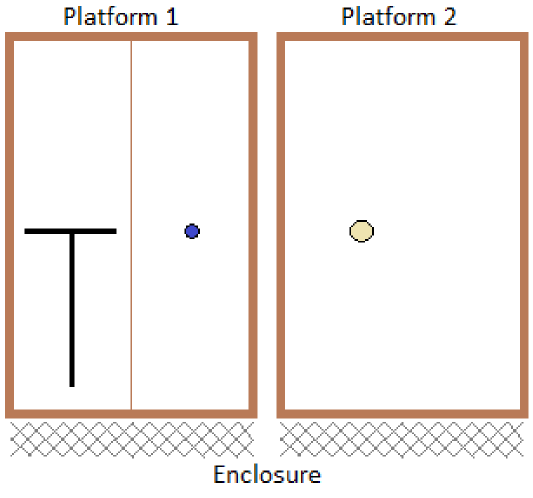
6.5.2 Experimental procedure and apparatus

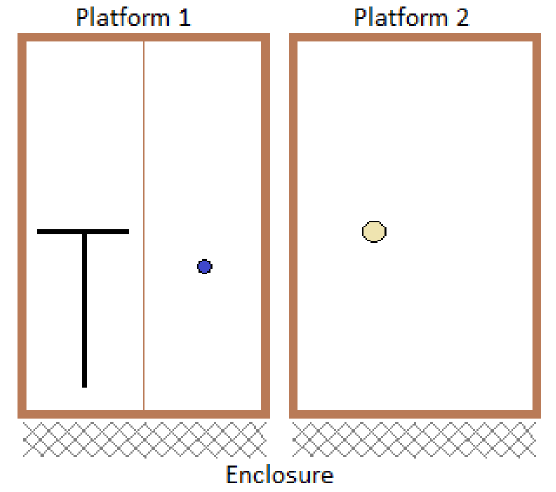
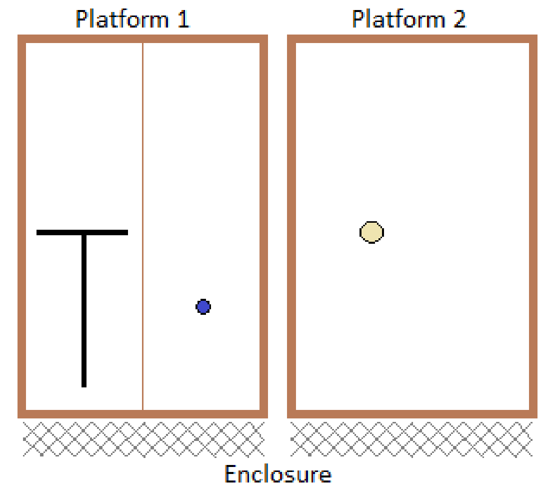
Experiments 2-4 presented a similar experimental set-up to that of Experiment 1, assessing if failure to select the rake during this experiment had been due to a lack of understanding of the task demands, which the reinforcement experiment aimed to correct, or a lack of self-control. During training and Experiment 1, the small reward had been placed at an equivalent distance as the rake handle. It is possible the equivalent distance of these items presented a difficult self-control challenge for subjects, once the large reward was moved to platform 2 during Experiment 1. Experiments 2-4 therefore presented an incremental movement of the small reward from out of reach in Experiment 2, to within reach during Experiment 3, to closer to the distance of the rake handle during Experiment 4. It was expected these experiments would present an increasing challenge to self-control as the small reward was moved closer to the subject. The gradual movement of the small reward was intended to slow down impulsive behaviour displayed during Experiment 1, assessing if subjects could develop self-control when presented with these incremental steps.

The experimental procedure and apparatus across Experiments 2-4 were as described for Experiment 1 (section 6.3.2), with the following exceptions (see Table 6.7). The distance of the small reward from the edge of platform 1 closest to the gibbon was 60cm during Experiment 2, out of gibbons' reach. This distance was reduced to 50cm in Experiment 3, which was within the gibbons' reach, and 40cm during Experiment 4. The rake remained at a reachable distance of 20cm across Experiments 2-4. If the small reward was selected, this was either passed to the subject during Experiment 2, or could be reached directly by the subject during Experiments 3 and 4 (Figure 6.7). If selected, the rake was immediately taken to platform 2, placed and baited within reach, allowing subjects to use the rake to retrieve the larger reward. A maximum of 120 trials were given during each experiment. Criterion level was classified as 8/10 selections of the rake, and therefore large reward, on two consecutive blocks.

All testing was recorded using a video camera for later behavioural coding. A note of the large reward position (left/right) and reward selected (large/small) was taken during testing for all experiments. The time to first contact, with either the rake or the small reward, was coded from the testing videos. Time to first contact was taken from the time when subjects entered the target area and attended to the apparatus until a first contact was made and aimed to measure subjects' impulsivity across the different experimental set-ups.

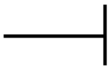
Table 6.7: Experimental set up for Experiments 2-4, showing the position of the small reward and functional rake on platform 1, and the large reward on platform 2. The rake handle is within gibbons reach across Experiments 2-4, with the large reward being out of reach on all experiments. The small reward is within reach on Experiments 3 and 4.

Experiment	**Experimental set-up**	**Apparatus placement**
Experiment 2		Platform 1: -Functional rake (20cm) -Small reward (60cm) Platform 2: -Large reward (60cm)

Experiment 3		Platform 1: -Functional rake (20cm) -Small reward (50cm) Platform 2: -Large reward (60cm)
Experiment 4		Platform 1: -Functional rake (20cm) -Small reward (40cm) Platform 2: -Large reward (60cm)

 = Large reward.

 = Small reward.

 = Functional rake.

6.5.3 Experiment 2 results

The small reward was positioned out of reach during Experiment 2, assessing if subjects would reliably select a functional rake and then use this to retrieve the larger reward, rather than selecting the small reward which was immediately passed to the subject if selected. All subjects reached criterion level performance within two blocks, selecting the rake (making the large reward accessible) on all trials (see Table 6.8).

Table 6.8: Subjects' genus, age group and sex. Number of selections of the large reward in each block completed in Experiment 2 (1 block = 10 trials), and across the total number of trials completed by each subject (significance on binomial test).

Subject	**Genus**	**Age group**	**Sex**	**Total number of large reward selections per block**		**Total number of large reward selections**
				1	**2**	
Betty	*Hoolock*	Adult	F	10	10	20/20**
Chan Thar	*Hoolock*	Adult	F	10	10	20/20**
Truman	*Hylobates*	Adult	M	10	10	20/20**
Violet	*Hylobates*	Sub-adult	F	10	10	20/20**
Asteriks	*Nomascus*	Adult	F	10	10	20/20**
Lucia	*Nomascus*	Sub-adult	F	10	10	20/20**

** Significant at $p < 0.01$ level.

6.5.3.1 Order effects

Order effects were assessed using Mann-Whitney U tests for enclosures 9 and 15, where two individuals participated in testing. Time to first contact over the first ten trials was compared. There was no significant difference between Violet and Truman ($U = 47.5$, $z = -0.19$, $p = 0.85$). There was a significant difference between Lucia and Asteriks ($U = 17$, $z = -2.5$, $p = 0.01$). Asteriks, the second individual tested, was faster to first contact (mean = 1.09 seconds) than Lucia (mean = 1.83 seconds), however, as all individuals selected the large reward on 10/10 trials, testing order was not considered to influence successful selection of the functional rake.

6.5.4 Experiment 3 results

During Experiment 3 the small reward was moved to within subjects reach, but still placed at a further distance from subjects than the rake handle. This experiment assessed whether subjects could refrain from immediately picking up the small reward, instead selecting the rake. Performance across all trials completed by each subject was evaluated using binomials tests. Three subjects selected the large reward significantly above chance across all trials and reached criterion level performance (see Table 6.9). The remaining three subjects failed to significantly select the large reward across all trials, or to reach criterion level, and so did not progress to Experiment 4 (see section 6.9 for further assessment with these subjects).

Table 6.9: Subjects' genus, age group and sex. Number of selections of the large reward in each block completed in Experiment 3 (1 block = 10 trials), across the total number of trials completed by each subject, and the two criterion blocks (significance on binomial test). Blocks where subjects reached criterion level performance are highlighted.

Subject	Genus	Age group	Sex	Total number of large reward selections per block												Total number of large reward selections	Performance on criterion blocks
				1	2	3	4	5	6	7	8	9	10	11	12		
Betty	*Hoolock*	Adult	F	5	7	3	1	1	2	4	5	4	8	5	7	52/120	n/a
Chan Thar	*Hoolock*	Adult	F	1	0	0	0	0	0	0	0	0	1	3	2	7/120	n/a
Truman	*Hylobates*	Adult	M	10	10											20/20**	20/20**
Violet	*Hylobates*	Sub-adult	F	9	10											19/20**	19/20**
Asteriks	*Nomascus*	Adult	F	6	2	2	3	3	3	3	6	6	4	3	5	46/120	n/a
Lucia	*Nomascus*	Sub-adult	F	7	9	9										25/30**	18/20**

■ = Criterion level performance.

** Significant at $p < 0.01$ level.

6.5.4.1 Order effects

Mann-Whitney U tests were used to assess order effects for those enclosures were two individuals were tested, with time to first contact over the first ten trials compared. A significant difference was found between both Violet and Truman ($U = 23.5$, $z = -2.01$, $p = 0.04$), and Asteriks and Lucia ($U = 13$, $z = -2.81$, $p < 0.01$). Truman, who was tested second, selected the large reward more frequently (10/10 trials) than Violet (9/10 trials) and he took longer to first contact (mean = 2.97 seconds, Violet mean = 2.18 seconds). This could suggest a learning advantage, however as both gibbons performed above chance with only one response separating them, the effects of testing order were minimal. Lucia was tested second in enclosure 15, and selected the large reward more frequently (7/10 trials) than Asteriks (6/10 trials). She was also faster to first contact (mean = 1.61 seconds, Asteriks mean = 2.72 seconds), suggesting she may have benefitted from witnessing Asteriks taking part in testing. However, as this was the only instance where the second tested individual both selected the functional rake more frequently and made first contact faster, and again, only one response separated both subjects, testing order did not appear to consistently improve performance.

6.5.5 Experiment 4 results

Following from Experiment 3, this experiment aimed to assess if those subjects who had reached criterion with the small reward within reach could continue to do so once this reward was moved closer towards them. Performance across all trials completed by each subject was evaluated using binomials tests. All three subjects selected the large reward above chance in all trials and reached criterion level, with two doing so in the minimum of two blocks (see Table 6.10).

Table 6.10: Subjects' genus, age group and sex. Number of selections of the large reward in each block completed in Experiment 4 (1 block = 10 trials), across the total number of trials completed by each subject, and the two criterion blocks (significance on binomial test). Blocks where subjects reached criterion level performance are highlighted.

Subject	**Genus**	**Age group**	**Sex**	**Total number of large reward selections per block**				**Total number of large reward selections**	**Performance on criterion blocks**
				1	**2**	**3**	**4**		
Truman	*Hylobates*	Adult	M	10	10			20/20**	20/20**
Violet	*Hylobates*	Sub-adult	F	8	3	10	8	29/40**	18/20**
Lucia	*Nomascus*	Sub-adult	F	8	9			17/20**	17/20**

= Criterion level performance.

** Significant at $p < 0.01$ level.

6.5.5.1 Order effects

Order effects were assessed using Mann-Whitney U tests for enclosure 9 where two individuals were tested, with time to first contact over the first ten trials compared. No significant difference was found between Violet and Truman ($U = 29$, $z = -1.59$, $p = 0.12$), again suggesting testing order did not influence performance.

6.6 Repeat of Experiment 1

6.6.1 Experimental procedure, apparatus and subjects

Successful subject's selections of the rake during Experiments 3 and 4, when the small reward was within reach, may have been based on the rule 'select whatever is closest'. Returning to Experiment 1 aimed to assess the three successful subject's capabilities when both the rake and small reward were again presented at an equal distance. This positioning potentially presented a more difficult self-control challenge, and the repeat of Experiment 1 assessed if the incremental steps given across Experiments 2-4 had helped subjects develop task-appropriate behaviour. The three subjects who reached criterion level during Experiment 4 were presented with an additional six blocks of trials during this repeat of Experiment 1 (see Table 6.11), with the experimental procedure and apparatus identical to that described in section 6.3.2.

Table 6.11: Subjects' species, age, sex, housing information, and whether they reached criterion level performance or not during the repeat presentation of Experiment 1.

Subject	Enclosure	Species	Age	Sex	Group	Exp. 1 repeat
Lucia	15	*N. leucogenys*	6	F	Family	✓
Truman	9	*H. pileatus*	13	M	Male/female pair	✓
Violet	9	*H. pileatus*	7	F	Male/female pair	✓

✓ = Reached criterion level.

6.6.2 Repeat of Experiment 1 results

All three subjects reached criterion level, taking between four and six blocks to do so (see Table 6.12), suggesting Experiments 2-4 had aided subjects in learning the task demands, with subjects displaying self-control when both the rake handle and small reward were placed at an equivalent distance. Performance across all trials completed by each subject was evaluated using binomials tests. Truman, who took the greatest number of blocks to reach criterion, was the only subject to select the large reward significantly above chance levels across all trials. Violet and Lucia performed poorly in the first two blocks before reaching criterion in blocks three and four.

Figure 6.6: Pileated gibbon (*H. pileatus*) successfully selecting the rake during the repeat of Experiment 1. The small reward remains in the starting position on the right of the testing platform.

Three gibbons were capable of inhibiting selection of an immediately attainable reward in order to select a rake for future reward retrieval, as has been found with large apes (Osvath & Osvath, 2008). It is possible gibbons required the experience gained during Experiments 2-4 in order to display task sufficient performance, with previous research finding self-control abilities improved across repeated testing of inhibitory control in task-experienced capuchins (*C. apella*) (Beran et al., 2016b). The progression of Experiments 2-4 was developed to address the possibility that subjects' failure during the first presentation of Experiment 1 was due to self-control issues, aiming to present a progressive difficulty in self-control as the reward was moved closer to subjects, allowing for assessment of at which point subjects failed to inhibit selection of the small reward. However, given that the three successful subjects all selected the rake higher than chance levels across Experiments 2-4, subjects may have performed at significant levels if given a repeat of Experiment 1 immediately following the reinforcement experiment. That is, their performance in Experiment 1 may have been due to a lack of understanding of the task demands, rather than a lack of self-control.

Table 6.12: Subjects' genus, age group and sex. Number of selections of the large reward in each block completed on the repeat presentation of Experiment 1 (1 block = 10 trials), across the total number of trials completed by each subject, and the two criterion blocks (significance on binomial test). Blocks where subjects reached criterion level performance are highlighted.

Subject	Genus	Age group	Sex	Total number of large reward selections per block						Total number of large reward selections	Performance on criterion blocks
				1	2	3	4	5	6		
Truman	*Hylobates*	Adult	M	7	6	4	7	10	8	42/60**	18/20**
Violet	*Hylobates*	Sub-adult	F	2	4	8	9			23/40	17/20**
Lucia	*Nomascus*	Sub-adult	F	3	4	9	9			25/40	18/20**

■ = Criterion level performance.

** Significant at $p < 0.01$ level.

6.6.2.1 Order effects

Mann-Whitney U tests were used to assess order effects for enclosure 9, where individuals may have witnessed each other participating in testing. Time to first contact across the first ten trials was compared, with no significant difference found between Truman and Violet ($U = 47.5$, $z = -0.19$, $p = 0.85$), suggesting testing order did not affect performance.

6.7 Comparison across Experiments 1-4

All subjects failed to significantly select the rake (leading to large reward retrieval) during Experiment 1, potentially due to either difficulty displaying self-control or a lack of understanding of the task demands. It was expected the progression over Experiments 2-4 and the repeat of Experiment 1 would present a more difficult self-control challenge as the small reward was moved towards an equivalent distance as the rake handle. There were individual differences in performance across the three subjects who completed all experiments; however, all subjects performed poorer on the repeat presentation of Experiment 1 than they had during Experiments 2-4, suggesting this experimental set-up did present a more challenging assessment of self-control. Violet and Lucia reached criterion level in fewer blocks than Truman during the repeat of Experiment 1; however, this was the only experiment where these two subjects failed to significantly select the functional rake across all trials. In contrast, Truman did select the functional rake at significant levels; however, this was the only experiment where he required more than the minimum of two blocks of trials to reach criterion level.

Considering subjects initial responses to the different experimental set-ups by comparing the first ten trials of each experiment (see Figure 6.7), the subjects selected the rake least frequently during the first presentation of Experiment 1, with both Lucia and Violet selecting the rake on 0/10 trials. Truman selected the rake on 5/10 trials during Experiment 1, suggesting this subject may have had some understanding of the task demands. Truman selected the rake on 10/10 trials during the initial trials of Experiments 2, 3 and 4, again suggesting this subject understood the task demands, with the improved performance on these experiments potentially being due to the easier self-control challenge than that presented during Experiment 1. Violet and Lucia also selected the rake on 10/10 of the initial trials of Experiment 2, suggesting the placement of the small reward out of reach during this experiment presented an easier self-control challenge for all three subjects, allowing them to make more considered responses and obtain the larger reward. However, both Lucia and Violet's selections of the rake dropped during the initial trials of Experiments 3 and 4, when the small reward was placed in an immediately retrieval position, leading to lower levels of self-control in these subjects. During the repeat of Experiment 1, all three subjects performed poorer than across Experiments 2-4, suggesting the equivalent distance of the rake handle and small reward resulted in an increased difficulty for displaying self-control, as expected. All subjects selected the rake more frequently during the repeat of Experiment 1 than they had during the initial presentation, suggesting Experiments 2-4 led to an understanding of the task demands and aided development of self-control.

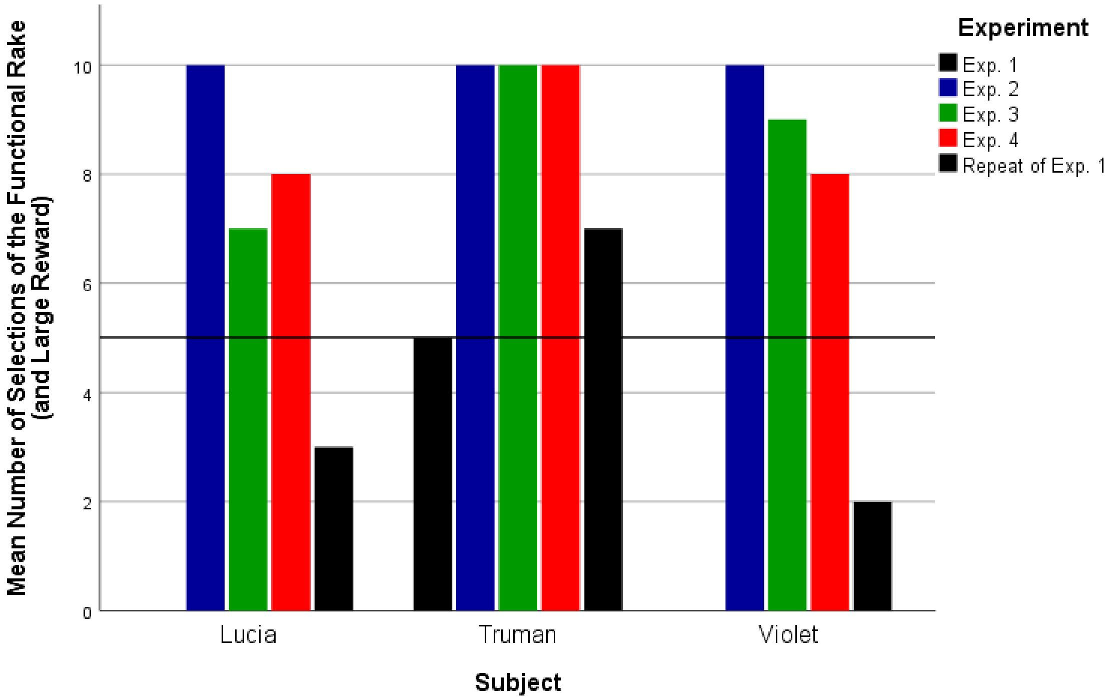


Figure 6.7: Mean number of selections of the functional rake useable to obtain the large reward during the first ten trials of Experiments 1-4, and the repeat presentation of Experiment 1, with chance level performance marked. Only subjects who completed all four experiments (n = 3) are included.

Time taken to first contact, with either the rake or small reward, during the first ten trials of each experiment did not show a consistent pattern in relation to the expected increase in difficulty of self-control requirements across experiments (see Figure 6.8). Truman, who selected the rake on all ten initial trials across Experiments 2-4, displayed the longest latency to first contact during these experiments, with the increased time spent considering the task before making first contact potentially resulting in more frequent selections of the rake (and large reward) by this subject.

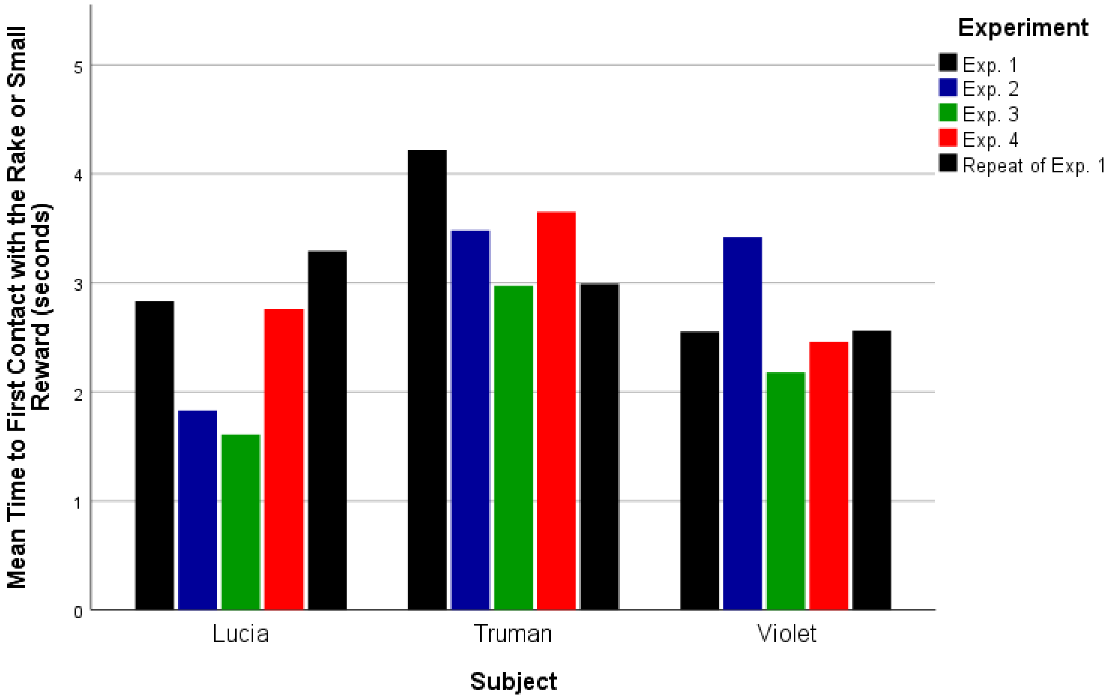


Figure 6.8: Mean time (seconds) to first contact with either the rake or small reward during the first ten trials of Experiments 1-4 and the repeat presentation of Experiment 1. Only subjects who completed all four experiments (n = 3) are included.

6.8 Further testing with successful subjects

The three subjects who successfully reached criterion performance on the repeat presentation of Experiment 1 participated in two further experiments. These experiments aimed to assess whether subjects selections during the previous experiments were reward directed, or whether selections were guided by a strong association with the functional rake or a rule based procedure of selecting whichever item was positioned closer to them.

6.8.1 Subjects

Subjects included two pileated (*H. pileatus*) and one northern white cheeked (*N. leucogenys*) gibbon (see Table 6.13). Subjects were separated and tested individually.

Table 6.13: Subjects' species, age, sex, housing information, and whether they reached criterion level performance or not across Experiments 5 and 6.

Subject	Enclosure	Species	Age	Sex	Group	Exp. 5	Exp. 6
Truman	9	*H. pileatus*	13	M	Male/female pair	✓	✓
Violet	9	*H. pileatus*	7	F	Male/female pair	✓	✓
Lucia	15	*N. leucogenys*	6	F	Family	✓	✓

*Criterion level performance.

6.8.2 Experiment 5

Subjects may have performed well on previous experiments due to a strong association with the functional rake, rather than selections being reward directed. Experiment 5 aimed to assess whether subjects would select a rake, which had led to large reward retrieval across previous experiments, or an immediately available large reward. If subjects selected the rake, this would suggest their behaviours were guided by a strong association with this functional rake, whereas if subjects selected the large reward, this would suggest their selections were reward directed. The experimental set-up and procedure were identical to Experiment 1 (see section 6.3.2), except the position of the rewards were reversed (see Figure 6.9). A small reward was placed out of reach (60cm) on testing platform 2, with a functional rake and large reward then placed on testing platform 1, both positioned within reach at 20cm from the platform edge. Subjects could select the large reward or rake by means of passing their hand over a line marked 20cm back from the edge of platform 1. After selection, the unselected item was removed from reach. If selected, the rake was immediately taken to platform 2 and baited within reach with the small reward. If subjects' previous selections had been guided by a strong association with the rake, it was expected selections of the rake would initially be high, selecting the rake instead of the immediately attainable large reward.

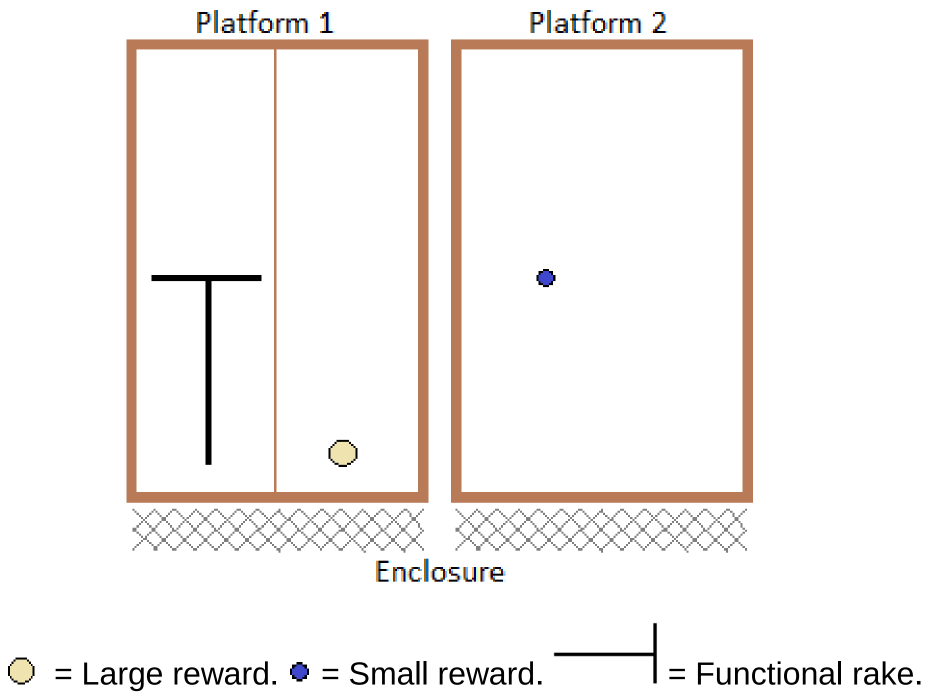


Figure 6.9: Experimental set up when gibbon made a choice, showing the functional rake and large reward positioned within reach at a distance of 20cm from the subject on testing platform 1, and the small reward placed out of reach at a distance of 60cm on platform 2.

Subjects were given a maximum of 60 trials (six blocks). Inter-trial intervals of 20 seconds and inter-block intervals of five minutes were enforced. Subjects were required to attain the large reward 8/10 times on two consecutive blocks in order to proceed to Experiment 6. The small reward used throughout was a blueberry, with the large being a piece of sweet potato. The position of the rake was randomised, appearing equally on both sides and never on the same side more than twice in succession. Reward position and whether subjects selected the rake or large reward was noted during testing.

Figure 6.10: Pileated gibbon (*H. pileatus*) looking to the large reward after selecting the rake during Experiment 5.

6.8.2.1 Experiment 5 results

All three subjects reached criterion level performance during Experiment 5 (see Table 6.14) suggesting subjects selections were reward directed, rather than being guided by an association with the rake which had led to large reward retrieval in previous experiments. Performance across all trials completed by each subject was evaluated using binomials tests, with all subjects selecting the large reward significantly above chance across all trials, rather than the rake useable to retrieve the small reward. Only Truman took more than minimum of two blocks to reach criterion level; however, this subject did not show a marked decrease in performance during the first block, performing close to criterion level.

Table 6.14: Number of selections of the large reward in each block completed in Experiment 5 (1 block = 10 trials), across the total number of trials completed by each subject, and the two criterion blocks (significance on binomial test). Blocks where subjects reached criterion level performance are highlighted.

Subject	Genus	Total number of large reward selections per block			Total number of large reward selections	Performance on criterion blocks
		1	2	3		
Truman	*Hylobates*	7	9	9	25/30**	18/20**
Violet	*Hylobates*	8	8		16/20*	16/20*
Lucia	*Nomascus*	9	10		19/20**	19/20**

= Criterion level performance.

* Significant at $p < 0.05$ level.

** Significant at $p < 0.01$ level.

6.8.3 Experiment 6

Subjects may have performed well on Experiments 2-4 by basing selections on the rule ‘select the closest object’, with Experiment 1 and Experiment 5 reflecting a more difficult decision between two equally distanced choices. The results of the repeat presentation of Experiment 1 and Experiment 5 suggest subjects behaviours were reward directed when the rake and reward were positioned at equivalent distances. Experiment 6 provided a further assessment of whether subjects’ selections were reward directed when one item was placed closer to the subject than the other, or whether subjects had been applying a rule-based system of selecting whichever item was positioned closest to them.

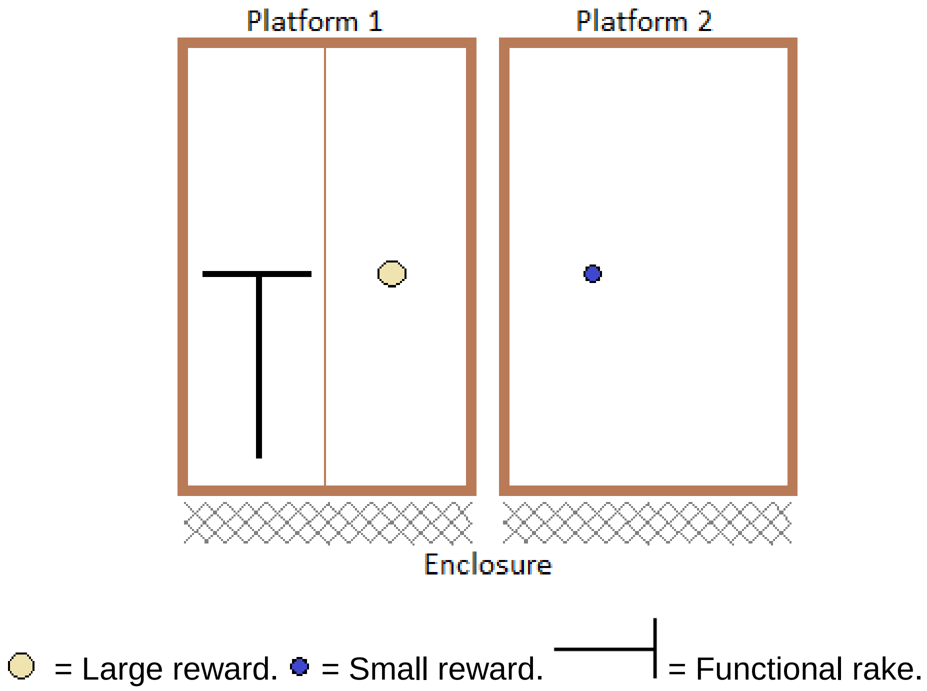


Figure 6.11: Experimental set up when gibbon made a choice, showing the functional rake and large reward positioned on platform 1, and the small reward positioned on platform 2. The rake handle is within reach at a distance of 20cm from the subject, with the large and small rewards placed out of reach at a distance of 60cm on platform 2.

Experimental procedure and apparatus were identical to Experiment 2 (see section 6.5.2), except the position of the rewards was reversed (see Figure 6.11). A functional rake was presented within reach (20cm), and a large reward placed out of reach (60cm) on testing platform 1, with a small reward placed out of reach (60cm) on platform 2, on the equivalent side as the rake. Thus, if subjects' selections were guided by the 'closeness' of items, it was expected they would select the rake. Alternatively, if subjects' behaviour was reward directed, it was expected they would select the out of reach large reward. A subject was classified as making a selection if their hand passed over a line marked 20cm back from the edge of platform 1. If the large reward was selected, by subjects reaching their hand over the 20cm line marked in front of the reward, it was passed to the subject. If the rake was selected, it was taken to platform 2 and baited with the small reward. The unselected item was removed from reach following selection. A maximum of 60 trials were presented, with inter-trial intervals of 20 seconds and inter-block intervals of five minutes. Criterion level was classified as 8/10 selections of the large reward over two consecutive blocks. The small reward was again a blueberry and the large reward a piece of sweet potato, with the position of the large reward being randomised across trials. Reward position and whether subjects selected the rake or large reward was noted during testing.

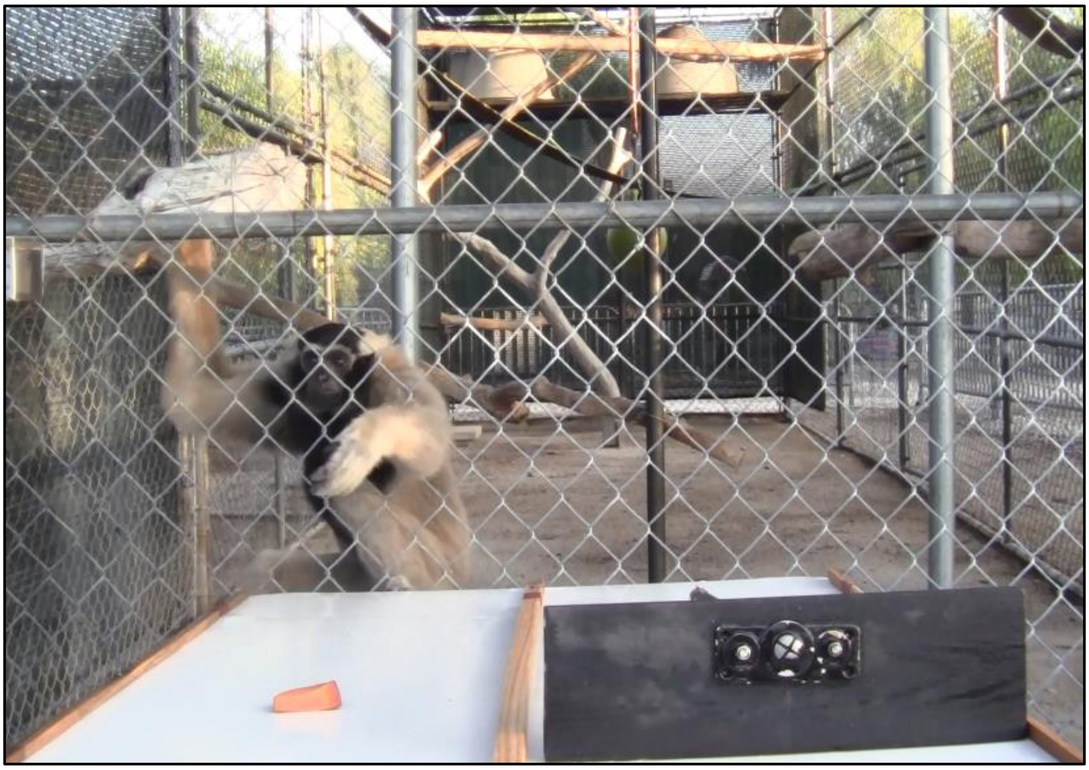


Figure 6.12: Pileated gibbon (*H. pileatus*) following selection of the large reward during Experiment 6.

6.8.3.1 Experiment 6 results

All three subjects reached criterion level during Experiment 6 (see Table 6.15); however, the number of blocks required to do so was greater than that during Experiment 5. Subjects may have initially been applying the rule ‘select whichever item is closer’ before learning to make selections guided by reward preference during this experiment. Performance across all trials completed by each subject was evaluated using binomials tests. Violet and Lucia, but not Truman, selected the large reward significantly above chance levels across all trials. However, performance of all subjects was significantly above chance on the last two blocks to reach criterion. These findings suggest subjects’ selections during Experiments 2-4 may also have been guided by the positioning of the rake and rewards on the testing platform.

Table 6.15: Number of selections of the large reward in each block completed in Experiment 6 (1 block = 10 trials), across the total number of trials completed by each subject, and the two criterion blocks (significance on binomial test). Blocks where subjects reached criterion level performance are highlighted.

Subject	**Genus**	**Total number of large reward selections per block**					**Total number of large reward selections**	**Performance on criterion blocks**
		1	**2**	**3**	**4**	**5**		
Truman	*Hylobates*	1	3	7	10	10	31/50	20/20**
Violet	*Hylobates*	1	7	9	10		27/40*	19/20**
Lucia	*Nomascus*	2	6	10	10		28/40*	20/20**

= Criterion level performance.

* Significant at p < 0.05 level.

** Significant at p < 0.01 level.

6.9 Subjects who failed Experiment 3

It is possible that the three subjects who failed Experiment 3 did not have a general preference for the large reward type (sweet potato) over the small reward type (blueberry), therefore making the increased effort required to obtain the large reward less beneficial. In order to ascertain the reward preference of these subjects, they completed ten trials presenting a choice between a piece of sweet potato or blueberry of equal size, without the presence of the rake. The rewards were placed out of reach (60cm) on opposite sides of a testing platform positioned outside subjects' enclosures, with reward position being randomised so that each reward appeared equally on each side of the platform but never more than twice in succession on one side. Following baiting, subjects were able to select one reward by means of passing their hand over a line marked 20cm back from the testing platform edge closest to the enclosure. Once selected, the chosen reward was passed to the subject with the non-chosen reward removed; a 20 second inter-trial interval took place and the next trial began.

Once subjects general reward preference was assessed, a re-assessment of large reward preference took place. The procedure was the same as during the general reward preference trials (see section 6.2.2), however, a small and large piece of the reward option predominantly chosen during the general reward preference stage was now presented (both small and large rewards were of the same type). Subjects were required to select the large reward eight or more times over ten trials to pass to a repeat presentation of Experiment 3.

Experimental apparatus and procedure during the repeat of Experiment 3 were identical to the initial presentation (see section 6.5); however, both the small and large rewards were now the reward type subjects preferentially selected during the reward preference trials. This presentation aimed to assess if subjects would refrain from selecting the immediately attainable reward, instead selecting the rake now that it led to a larger piece of the same reward. The maximum number of trials was limited to 60 (six blocks) per subject. All testing was recorded with a video camera, with a note of reward positions and which reward was selected (potato/blueberry then small/large) taken during testing. The time to first contact with either the rake or the small reward was coded from the testing videos for the repeat presentation of Experiment 3. This was taken from the time when subjects entered the target area and attended to the apparatus until first contact was made.

6.9.1 Reward preference trials

Betty and Asteriks preferentially chose the potato during the general reward preference trials, however, they both selected this option on only 6/10 trials (see Table 6.16). These subjects lack of strong preference for the sweet potato may have affected performance during Experiment 3, where the greater effort incurred to select the rake and then use this to retrieve a piece of potato may have been outweighed by an immediately available blueberry. Chan Thar selected the blueberry on 8/10 trials. Again, this subject’s performance during Experiment 3 may therefore have been guided by reward preferences, where her preferred reward (blueberry) was immediately available. All three subjects selected the large reward over the smaller option on 9/10 reward size preference trials. Given Chan Thar’s preference for blueberries, she was presented with a small reward of one blueberry and a large reward of three blueberries during the repeat of Experiment 3. Betty and Asteriks did not show a strong preference for either reward type, but did preferentially select a larger reward. These two subjects were therefore tested with small and large pieces of potato during the repeat of Experiment 3, as they displayed a slight preference for this food.

Table 6.16: Subjects' species, age, sex and housing information. The most frequently selected reward during the ten general reward preference assessment and number of times the large reward was selected during the ten large reward preference trials.

Subject	Enclosure	Species	Age	Sex	Group	General reward preference	Large reward preference
Betty	10	*H. leuconedys*	17	F	Male/female pair	Potato 6	9
Chan Thar	6	*H. leuconedys*	10	F	Male/female pair	Blueberry 8	9
Asteriks	15	*N. leucogenys*	16	F	Family	Potato 6	9

6.9.2 Repeat of Experiment 3 results

When using small and large options of the same reward type, Betty now reached criterion within two blocks, preferentially selecting the larger option (see Table 6.17). Asteriks and Chan Thar failed to reliably select the larger option, with Chan Thar developing a side bias over five of the six presented blocks. Performance across all trials completed by each subject was evaluated using binomials tests, with only Betty selecting the large reward significantly above chance levels.

Table 6.17: Number of selections of the large reward in each block completed by subjects (1 block = 10 trials), across the total number of trials completed by each subject, and the two criterion blocks (significance on binomial test). Blocks where subjects displayed a side bias or reached criterion level performance are highlighted.

Subject	**Genus**	**Total number of large reward selections per block**						**Total number of large reward selections**	**Performance on criterion blocks**
		1	**2**	**3**	**4**	**5**	**6**		
Betty	*Hoolock*	8	8					16/20*	16/20*
Chan Thar	*Hoolock*	4	5	5	5	5	5	29/60	n/a
Asteriks	*Nomascus*	7	7	6	3	3	6	32/60	n/a

= Side bias.

= Criterion level performance.

* Significant at $p < 0.05$ level.

6.10 Discussion

This series of experiments found a small number of gibbons were capable of displaying self-control, foregoing an immediate small reward and selecting a functional rake in order to retrieve a larger reward, as has been found in large apes (Osvath & Osvath, 2008). In line with previous assessments of self-control in primates (e.g., Anderson et al., 2010; Beran et al., 2016b; Evans & Beran, 2007), individual differences in performance were found; three individuals successfully selected the functional rake when this was placed at an equal distance as a small reward, with the remaining three subjects failing to select the rake once a small reward was within reach. Given these notable individual differences, and the limited number of individuals tested, further assessment of these capabilities is needed; however, these experiments present a preliminary investigation of self-control within the small apes.

All subjects initially failed to select the functional rake (leading to large reward retrieval) during Experiment 1. Gibbons' poor performance was potentially due to difficulties in understanding the task demands, failing to relate the functional rake to retrieval of the large reward on the second platform. The reinforcement experiment aimed to make this relationship clear to the subjects, with all subjects significantly selecting the rake (and large reward) during this experiment. Experiments 2-4 aimed to assess whether subjects' failure during Experiment 1 had been due to a lack of understanding of the task demands or a lack of self-control. The small reward was moved incrementally during these experiments, from out of reach and so not immediately attainable, to closer in distance to the subject, and to the distance of the rake handle. All subjects selected the rake during Experiment 2, with the position of the small reward out of reach potentially making this an easier assessment of self-control. Three subjects failed as soon as the small reward came within reach during Experiment 3, suggesting an inability to suppress the impulse to reach for the food item that was in reach, despite it being the smaller reward.

The three subjects who were successful across Experiments 2-4 and the repeat of Experiment 1 showed varying performance, both in frequency of rake selections and latency to first contact with either the rake or small reward.

However, all three reliably displayed self-control and reached criterion level across experiments, displaying comparable abilities to large apes and New World monkeys by inhibiting consumption of more immediate rewards in order to access delayed larger rewards (Evans & Westergaard, 2006; Osvath & Osvath, 2008). The task used by Evans and Westergaard (2006) required subjects to continually refrain from eating the tool until the large reward had been retrieved, in contrast to current experiments where the temptation of the small reward was removed immediately following selection of the rake. Although tasks such as Evans and Westergaard's (2006), requiring sustained inhibition, may appear to present more difficult self-control challenges, previous assessments with capuchins (*C. apella*) have found little difference in performance across self-control tasks requiring different levels of inhibition (Beran et al., 2016b).

The current task was comparable to that presented by Osvath and Osvath (2008), with both large apes and three of the gibbons tested reliably selecting a functional object for large reward retrieval rather than an immediately attainable small reward. However, the level of self-control displayed by gibbons was lower than that found in large apes and did not confirm subjects mentally represented the reward retrieval event. During Osvath and Osvath's (2008) task, subjects were required to retain and transport the tool to a reward retrieval area, which was not visible during tool selection. A delay period of 70 minutes was imposed between tool selection and reward retrieval, and a total of 14 trials were given to each subject. In contrast, gibbons had extensive experience during the current experiments, requiring progressive training steps in order to successfully select the functional tool when presented at an equivalent distance as a small reward. Furthermore, subjects endured a minimal delay between selection and retrieval periods, and were not required to transport the rake, keeping in mind the future reward retrieval over an extended delay period.

Performance levels of the three successful subjects during the second presentation of Experiment 1 were improved from the first assessment. The presentation of the rake handle and small reward at equivalent distances during this experiment appeared to present a more difficult self-control challenge, as subjects displayed a poorer performance during the repeat of Experiment 1 than they had across Experiments 2-4, either taking longer to reach criterion level or failing to significantly select the rake across all trials. In line with Beran et al.'s (2016b) finding that capuchins (*C. apella*) experienced in cognitive testing displayed increased levels of inhibition following repeated experience of self-control tasks, gibbons improved performance may have resulted from increased experience of displaying self-control across Experiments 2-4. Alternatively, this improved performance may have been due to an initial lack of understanding of the task demands during Experiment 1. The three successful individuals selected the rake at highly significant levels across Experiments 2-4, and it is possible these subjects would have succeeded with the experimental set-up of Experiment 1 immediately following the reinforcement experiment.

It was expected that the incremental movement of the small reward across Experiments 2-4 would result in a gradual decrease in selections of the rake due to difficulty in displaying self-control; however, this was not found. It is possible the three successful subjects were applying the rule 'select whichever item is closest' during Experiments 2-4, where the rake handle was positioned closer than the small reward. This application of this rule based procedure may have contributed to the lack of gradual decrease in rake selections as the reward was moved closer to the subject over the experiments. That subjects may have been selecting whichever item is placed closer is supported by the findings of Experiment 6, where subjects initially selected the rake when this was presented at a closer distance than a large reward, before gradually overcoming this bias and selecting the large reward.

The three subjects that failed to select the rake during Experiment 3 were further assessed with this experimental set-up following reward preference trials and confirmation that these subjects selected a large piece of their preferred reward over a small piece. This repeat of Experiment 3 aimed to assess whether subject's poor performance was due to an inability to display self-control or, alternatively, due to a preference for the smaller reward type guiding their selections. One subject, Chan Thar, did preferentially select blueberries during the reward preference trials. However, this subject again failed to reach criterion when presented with a small and large quantity of her preferred reward during the repeat of Experiment 3, suggesting her selections had not been guided by a preference for blueberries during the initial presentation. It is possible Chan Thar lacked the ability to display self-control once the small reward was placed within reach. However, this subject developed a side bias during the second presentation, resulting in selection of the rake on half of the trials. Given that she had not displayed this bias across previous experiments, suggesting she did understand the task demands, this may be reflective of decreasing interest in the task, leading her to repeatedly select from the same side of the platform.

The remaining two subjects, Asteriks and Betty, showed only a slight preference for potato during the reward preference trials, suggesting their selections during Experiment 3 were not guided by a preference for the small reward type, blueberries. Asteriks again failed to reach criterion level during the repeat of Experiment 3 with small and large rewards of sweet potato, suggesting this subject's poor performance was due to an inability to inhibit selection of the immediately attainable small reward. Betty reached criterion level during the repeat presentation of Experiment 3, when presented with small and large pieces of potato. It is surprising that Betty successfully refrained from selecting a small piece of potato, in order to select the rake and retrieve a large piece of potato, but did not do so during the initial Experiment 3, where the small reward was the less preferred blueberries. Despite failing to reach criterion level, Betty selected the rake more frequently during the second half of the 120 trials given in the first presentation of Experiment 3. It is possible the extended experience of the task led to the improved performance displayed by this subject during the repeat of Experiment 3. It is also of note that by presenting small and large quantities of the same reward type, the repeat presentation of Experiment 3 did not put conflicting desires in opposition. This may have presented an easier choice between rewards differing in quantity, rather than quantity and quality, leading to Betty's improved performance during the repeat presentation of this experiment.

Gibbons selections during Experiments 1-4 of the current research may have been guided by previously learnt associations between the functional rake and a food reward (see Chapters 3, 4 and 5). Findings from Experiment 5 suggested subjects' selections were reward directed, rather than based on a strong association with the functional rake. The three successful individuals selected the large reward both when this did, and did not require selection of the rake. However, the presence of the large reward on the second platform during the selection period of Experiments 1-4 likely served as a behavioural cue for subjects, with a desire for this visible reward guiding selections, rather than a mental representation of future reward retrieval. As noted by Osvath and Osvath (2008), the learning of arbitrary relations between stimulus and response need not suggest foresight of future reward retrieval events. Thus, although some gibbons displayed self-control, inhibiting selection of an immediately attainable small reward in order to obtain a larger reward, the current findings do not confirm gibbons acted with foresight, mentally representing future reward retrieval. Self-control, with or without a sense of the future, is an integral part of prospective cognition, without which current needs cannot be inhibited in favour of future desires. Further research is needed to assess whether gibbons can mentally represent a future event without cueing from current stimuli; however, the current findings present preliminary evidence for self-control capacities within these apes.

Chapter 7: General Discussion

7.1 Introduction

This thesis first presented an assessment of the ability to attend to functionally relevant features of objects in gibbons (*Hylobatidae*), a much understudied but diverse group of apes representing the only surviving divergence between the large apes and monkeys. An initial investigation into components of prospective cognition was then presented, including assessment of basic prospection and self-control abilities within the small apes. Chapter 3 reported most gibbons failed to attend to the functionally relevant features of two rakes during a raking-in task requiring consideration of the three dimensional relationship between the rakes and rewards. This failure was potentially a result of similarity in appearance between the rakes obscuring the functional differences between them, rather than a lack of ability to attend to the relevant features per se. Chapter 4 addressed this issue, presenting a series of experiments with more distinct perceptual differences between rakes. Eight gibbons were found to reliably select functional rakes, transferring knowledge of functionally relevant features across rake sets.

Chapter 5 presented an assessment of basic prospective cognition, finding gibbons were capable of selecting un-baited rakes for future use, connecting temporally and spatially distinct events of rake selection and reward retrieval. Finally, Chapter 6 presented an assessment of self-control, an integral component of prospective cognition without which individuals could not refrain from acting on current desires in order to benefit in future situations. Due to the limited number of subjects, the results from Chapter 6 are tentative; however, three gibbons displayed the ability to forego an immediate small reward, instead selecting a rake in order to retrieve a larger reward at an alternative location. An evaluation of these main findings will be covered in this chapter, including discussion of the components of prospective cognition assessed, and the applicability of learned associations accounting for subjects' behaviour. Differences in performance between genera, age groups and sexes will then be discussed. Finally, consideration of directions for future research will be given along with concluding remarks.

7.2 Evaluation of main findings

7.2.1 Attending to functionality

Previous research has found gibbons succeed during zero-order manipulation tasks, requiring them to draw in a rake shaped object in order to retrieve a food reward (Cunningham, 2006; Cunningham, Anderson & Mootnick, 2006, 2011; Inoue & Inoue, 2002). However, these assessments do not provide evidence of whether gibbons understand which features of the rakes make them functional for reward retrieval. Chapters 3 and 4 therefore presented a raking in task with one functional and one non-functional rake, assessing if gibbons could differentiate between these rakes, attending to the relevant features in order to successfully retrieve a food reward. Understanding which properties are relevant when considering the functionality of potential objects is crucial for successful tool use. Although there is little support for true tool use in gibbons (see Cunningham, 2006), the ability to consider appropriate features of objects may be beneficial during their primary mode of locomotion, brachiation. All apes are capable of brachiating, but the small apes have optimised this mode of locomotion. Gibbons utilise 'ricochetal brachiation', defined by Reichard, Barelli, Hirai and Nowak (2016) as locomotion where the body does not continuously contact the substrate during suspensory hand over hand movements, in order to reduce energy expenditure during fast arboreal travel. This movement requires quick reaction times and highly developed perception of both distance and suitability of potential branches or other substrates to which the individual is moving, with Carpenter (1940) suggesting gibbons may surpass the larger apes in visual perception speed.

Given the limited support for tool use in gibbons, the task used throughout the current research did not require true tool use, in that subjects were not required to re-orient the rakes before use (see Shumaker, Walkup &

Beck, 2011). However, this was considered an appropriate task in order to assess gibbons' ability to attend to the functionality of the presented rakes. Both tool using and typically non-tool using species of monkeys have been found to differentiate between functionally relevant and irrelevant features of cane shaped tools during raking-in tasks (Cummins-Sebree & Fragaszy, 2005; Fujita, Kuroshima & Asai, 2003; Hauser, 1997; Hauser, Pearson & Seelig, 2002), ignoring changes in irrelevant perceptual features such as colour, but attending to perceptual changes which impact on functionality, such as shape. These findings led Spaulding and Hauser (2005) to propose even non-tool using primates may be equipped with an innate mechanism which recognises functionally relevant features of objects, with appropriate experience needed for this mechanism to function efficiently. In line with this proposal, it was expected gibbons, given their phylogenetic position, would attend to the features of the rakes which impacted on functionality for reward retrieval.

The findings of pilot testing (Chapter 3) did not support this hypothesis; only one of the 23 subjects reliably selected the functional rake during an initial 60 trials. Despite the frequency of functional rake selections increasing slightly during the last 20 trials, performance remained close to chance levels for almost all subjects. As testing progressed, subjects spent less time in the target area surrounding the apparatus and were faster to first productive contact with the rakes. Given that performance did not greatly improve, this may be reflective of decreasing interest in the task and more impulsive behaviour. Multiple individuals developed side biases, suggesting difficulty in learning which aspects of the rakes to attend to, with subjects frequently selecting the rake positioned on the side of the testing platform they approached from. These findings are in line with the Anderson (2012), who suggested gibbons (*Hylobates leucogenys*, n = 3, *Symphalangus syndactylus*, n = 2) reverted to a side bias when they were frustrated or unmotivated during object permanence tasks.

The best performing subjects were given an additional 60 trials; however, this increased experience with the task resulted in only two of the six individuals given further trials reaching criterion level. The three successful individuals spent less time engaged with the task than unsuccessful subjects, being faster to first productive rake contact and spending less time in the target area surrounding the apparatus, suggesting more focussed and efficient manipulations. These subjects did not display a side bias during any block; however, it is unclear whether this was due to better understanding of the task or, conversely, is something which aided in understanding task requirements. Given their improved performance, it is possible the decreased time spent engaged with the task was a result of more efficient behaviour due to a better understanding of the task demands in these three individuals. Overall, and in contrast to findings with New World monkeys, the results of pilot testing provided little support for the ability to attend to the relevant perceptual features of the rakes in gibbons.

The task used during pilot testing was similar to the ‘inverted rake task’ developed by Povinelli and Reaux (2000), requiring consideration of the three dimensional relationship between the rakes and rewards. Assessments of spontaneous understanding of this task with chimpanzees (*Pan troglodytes*), vervet monkeys (*Cercopithecus aethiops*) and tamarins (*Saguinus oedipus*) found these species perform around chance levels (Povinelli & Reaux, 2000; Santos, Pearson, Spaepen, Tsao & Hauser, 2006). Povinelli and Reaux (2000) considered subjects may perceive the relationship between rakes and reward during the inverted rake task in a gestalt manner, considering whether the reward is positioned in front of a rake or not, rather than the causal interaction of the rake coming in to contact with the reward. They further suggest that an appreciation of the need for contact between the rake and reward may require an understanding of transfer of force, as opposed to lower level perceptual information processing required by considering ‘in front of or not’. These suggestions are supported by Santos et al.’s (2006) finding that tamarins (*S. oedipus*) and vervet monkeys (*C. aethiops*) performed poorer on an inverted rake task than a task with cane shaped tools where the reward was either positioned inside or outside the canes hook.

Reports of New World monkeys differentiating between functionally relevant and irrelevant features of tools have often presented tasks where the two dimensional relationship between tools and reward may be used to guide selections, with rewards either placed inside or outside of a hook at the end of a cane (Cummins-Sebree & Fragaszy, 2005; Fujita et al., 2003; Hauser, 1997; Hauser et al., 2002). Thus, differences in the ability to attend to the functionally relevant features of rakes between New World monkeys and gibbons during pilot testing may be due to task differences, rather than representing a cognitive deficit in gibbons.

When comparing gibbons' performance during pilot testing to the aforementioned species assessed with the inverted rake task, it was expected that, in line with Spaulding and Hauser's (2005) proposal, the extended testing period given to gibbons, allowing for greater experience with the task, would lead to improved performance. Chimpanzees selected the functional rake on 57% of the four presented inverted rake task trials with vervet monkeys succeeding on 60%, and tamarins 52%, of six trials presented to these two species (Povinelli & Reaux, 2000; Santos et al., 2006). Considering only the first five trials presented, as a group gibbons selected the functional rake on 57% of trials, a comparable performance to the previous species assessed. However, the extended testing period did not lead to improved performance, with the three successful gibbons selecting the functional rake at comparable levels to those species assessed with minimal task exposure. Win Bo, who reached criterion within 60 trials, and Lucia who required 100 trials to reach criterion, both selected the functional rake on 58% of trials. Truman, who performed best over all trials, selected the functional rake on 63% of 110 trials. These subjects learned to differentiate between the rakes, potentially attending to perceptual features such as differences in rake-end as shape. It is possible the similarity in appearance between the rakes, taking into account gibbons' direction of approach to task, obscured differences between them, leading to the generally low performance levels.

Chapter 4 addressed this issue, presenting more visually distinctive rakes in order to reduce the cognitive demands of the task by making differences more obvious. Furthermore, this chapter presented experiments where the rewards position in relation to the rake could be used to guide selections. In line with Povinelli and Reaux's (2000) suggestions that consideration of the two dimensional relationship between rake and reward may present an easier task than consideration of physical contact between objects, it was expected gibbons would display an improved performance on these experiments. All 18 subjects reached criterion level performance during Experiment 1, presenting a cross shaped non-functional and rectangular functional rake of differing colours, suggesting gibbons successfully differentiated between the rakes, attending to the more salient differences in appearance. Experiment 2 removed the colour differences between the rakes, encouraging subjects to attend to differences in rake shape or functionality. Of the 18 subjects, 15 reached criterion level, suggesting, as has been found for capuchins (*Cebus apella*) (Fujita et al., 2003), these subjects' selections were not tied to the more salient perceptual colour differences. Three subjects failed to significantly select the functional rake, with these subjects potentially learning to attend to colour differences in Experiment 1, and now struggling to re-learn which aspects of the rakes to attend to.

Experiments 3 and 4 manipulated rake shape, assessing if subjects could transfer knowledge of functionality across visually different, but functionally similar, rakes. When presented with the same rake set used during pilot testing, ten of the 15 subjects reached criterion level, with eight of these individuals also reaching criterion level during Experiment 4, which presented the cross shaped non-functional rake and a new smaller functional rake. These findings suggested successful subjects' selections were also not tied to specific rake shapes, with subjects transferring knowledge across rake sets. Several subjects failed to reach criterion level during Experiment 3 and 4, again suggesting these subjects had learned to attend to specific perceptual differences, failing to re-learn which aspects of the rakes to attend to within the limit of 30 trials.

During Experiments 1, 2 and 4 of Chapter 4, the reward was positioned off-set to the cross shaped non-functional rake. Gibbons selections may therefore have been guided by the rewards position, either in front of the rake or not. However, this cannot account for subjects' responses during Experiment 3, where the reward was positioned in front of both the functional and non-functional rake. As gibbons were successful with the experimental set-up used during Experiment 3 (Chapter 4), but most subjects failed this task during pilot testing, the experience gained from Experiments 1 and 2 may have aided subjects in attending to functionally relevant features during this more difficult task. Whether this extended experience led gibbons to consider the physical contact between the rake and food reward, or whether repeated testing with the same functional rake across Experiments 1-3 led subjects to select this rake is not clear from the current data. However, the findings of Chapter 4 support Spaulding and Hauser's (2005) proposal that, given task experience, non-tool using species such as gibbons can recognise the functionally relevant features of objects, with subjects discriminating between rakes in order to successfully retrieve a food reward. This is not to say these primates possess an innate understanding of functionality, but rather they are capable of differentiating between the features of functional and non-functional rakes.

Assessment of differences in performance, time spent attending to the task, and latency to first productive contact between the last trials of each experiment and the initial trials of the following trials suggested subjects were re-learning which aspects of the rakes to attend to during each experiment. Frequency of functional rake selections showed a general trend for being higher during the last ten trials of each experiment than the initial trials of the following experiment, suggesting a lack of spontaneous understanding of the task demands in some subjects as the changing dimensions affected performance. The time subjects spent attending to the task, and time to first productive rake contact was greater during the first ten trials of Experiments 2 and 4, than in the last trials of Experiments 1 and 3, again suggesting the changing dimensions resulted in more consideration of the task before selections were made. In contrast, time spent attending to the apparatus and to first productive contact decreased during the first ten trials of Experiment 3, which presented the pilot testing rake set, when compared to the last ten trials of Experiment 2. As performance was not improved during these initial trials, it is possible the previous experience with this rake set resulted in a form of learned helplessness; their earlier failure with these rakes reduced interest as they had learned they did not obtain the reward on most trials.

These performance differences between experiments decreased as testing progressed. It is possible subjects were 'learning to learn', that is, subjects viewed each experiment as a novel task but became better at discriminating between the rakes as testing progressed (Harlow, 1949; Tomasello & Call, 1997). However, given findings that gibbons struggle with learning set formation (see D'Agostino & Cunningham, 2015; Rumbaugh & McCormack, 1967), it is also possible subjects transferred knowledge across rake sets once they had learned which features to attend to, with the drop in performance levels reflective of the decreased perceptual features available to attend to in Experiment 2, and the introduction of one new rake shape in both Experiments 3 and 4, requiring subjects to learn which aspects of this novel rake to attend to. This contrasts with Fujita et al. (2003); finding capuchins (*C. apella*) maintained performance levels across experiments presenting novel shaped tools. It is possible differences in temperament between the species, or the increased tool using competency displayed by capuchins (see Shumaker, Walkup & Beck, 2011), resulted in faster learning and attentiveness to functionally relevant features of the tools. This suggestion is supported by Cummins-Sebree and Fragaszy's (2005) finding that capuchins displayed more exploratory manipulations during a raking in task than tamarins (*S. oedipus*), for which there are a smaller reported repertoire of tool using behaviours (Shumaker et al., 2011).

Despite several subjects successfully learning to attend to the relevant features of the rakes, overall, gibbons performed poorer than may be expected given their position within the primate lineage. As noted, gibbons are not prolific tool users, the few reports of wild tool use consist of these apes breaking off and dropping branches in the direction of human observers, using pendulum motions on vines or slender trees in order to reduce distances during locomotion, and pulling in vines to draw in out of reach food sources (Carpenter, 1940; Shumaker et al., 2011). The natural habitat and foraging patterns of gibbons present the ecological opportunity for tool use and goal directed object manipulation, but they do not require it. Furthermore, the sociality of gibbons is not overly conducive to the maintenance of tool use behaviour. Social learning is considered to be more efficient than independent learning, allowing individuals to focus on relevant information and acquire skills without requiring time consuming and potential dangerous innovation (Schuppli et al., 2017). Population differences in sociability and opportunities for social learning have been found to affect exploratory tendency in groups of orangutans (*Pongo abelii*, *Pongo pygmaeus wurmbii*) (Schuppli et al., 2017). Gibbons tend to live in small, monogamous family groups. The potential for social transmission of complex manipulation skills is therefore reduced to transmission by offspring dispersing to form new family groups. Thus, although non-tool using species may also possess an ability to attend to the functionally relevant features of objects, socio-ecology and species typical repertoires of behaviour may predict the form and flexibility of object manipulations and tool use (Cummins-Sebree & Fragaszy, 2005).

7.2.2 Prospective cognition

Encephalisation quotients (EQ), a measure of extra neural tissue available for cognitive processes when the effect of body size is removed, have traditionally been referred to when considering differences in cognitive ability between species (see Jerison, 1973). EQ is determined by calculating the ratio of a species absolute brain mass to its expected brain mass, and therefore represents whether a species brain is larger or smaller than would be expected for its body mass (Boddy et al., 2012). Using this method, gibbons have higher than expected relative brain sizes, as do all apes; however, such a viewpoint may assume larger bodied animals brains are simply 'scaled-up' versions of smaller bodied species brains, whereas anthropoid brains have been found to be both quantitatively and qualitatively different (Boddy et al., 2012; Rilling, 2006; Willemet, 2013). Deaner, Isler, Burkart and van Schaik (2007, see also, Shultz & Dunbar, 2010) found total brain size, rather than EQ, better predicted variation in cognitive ability in primates. Gibbon's absolute brain volume is smaller than all species of large apes, but larger than most species of monkey (Rilling, 2006). Despite this decreased brain volume when compared to the large apes, similarities have been found in the relative proportions of some brain areas between the small and large apes.

MacLeod, Zilles, Schleicher, Rilling and Gibson (2003) found that gibbons clearly group with the other apes, as opposed to monkeys, on cerebellar proportions. Although early reports considered the cerebellum as primarily contributing to motor control, more recent evidence suggests this area, which shows a marked increase in hominoids, is active in various higher cognitive functions including; movement planning, some forms of learning, problem solving, and attention shifting (Buckner, 2013; MacLeod et al., 2003). Considering the neocortex, proposed to be the area of higher cognitive functions and critical for learning and problem solving, similarities have also been found between gibbons and the large apes (Shultz & Dunbar, 2010). Rilling and Insel (1999) found apes tended to have both a larger neocortex volume than would be predicted by their body size, and a larger relative neocortex volume than the monkey species assessed, with the exception of capuchins (*C. apella*). Overall, gibbons appear to share more similarities in neuroanatomy with the large apes than with most monkeys, with larger relative neocortex and cerebellum volumes than predicted by their body mass. Areas of the neocortex, the temporal and parietal lobes in particular, have been proposed to be the site of a brain network common to forms of flexible self-projection, including episodic memory, theory of mind, navigation and future thinking (Buckner & Carroll, 2007).

Gibbons have high spatial and temporal variability in their diet, likely requiring spatial memory in order to remember both the location and content of food sources throughout their territory when navigating (Scheumann & Call, 2006). Asensio, Brockelman, Malaivijitnond and Reichard (2011) report gibbons travel paths are goal oriented and directed by foraging, with the subsequent location planned before leaving a food source (see section 2.3.1 for more details). Furthermore, Reichard and Sommer (1997) report many inter-group encounters by wild lar gibbons (*Hylobates lar*) occurred in or near preferential feeding sites, suggesting travel was directed by optimal foraging routes. The ability to keep in mind resources which are not currently perceivable may aid gibbons in creating spatial cognitive maps of their home ranges, allowing them to navigate effectively. There is some evidence for the presence of spatial memory in gibbons tested in captivity, with Scheumann and Call (2006) finding a southern yellow-cheeked gibbon (*Nomascus gabriellae*) displayed knowledge of what food item was hidden where during a foraging task simulating natural conditions. Tests of object permanence in captive gibbons provide further support for the ability to mentally represent out of sight objects (e.g., Anderson, 2012; Fedor, Skollár, Szerencsy & Ujhelyi, 2008). The requirement to keep track of territory boundaries, resource locations and optimal foraging routes present prime situations where future directed behaviour may be beneficial in natural contexts.

There is ongoing debate as to the behavioural criteria necessary for assessing future thinking in non-human animals within captive settings, with

Suddendorf and Corballis (1997; 2010) arguing only humans are capable of fully fledged mental time travel, projecting the mental self both forwards and backwards along a subjective time continuum. Focus on restrictive criteria, and assessment of future thinking similar to that displayed by humans, may overlook the subtle difference in prospective cognition between species. A broader focus, investigating component processes of prospective cognition and species specific adaptations allows for comparison across species, investigating the evolutionary emergence of these capacities. The current research aligns with Raby and Clayton's (2009) framework of prospective cognition, differentiating between complex future-oriented behaviours requiring a sense of the future, and basic prospective behaviours that do not. Fixed action patterns and learned associations are considered to be basic prospection; behaviours that do not require a sense of the future despite having future consequences. In contrast, complex prospection, including prospective memory, semantic and episodic future thinking, does require a sense of the future. Throughout the current thesis gibbons displayed basic prospective behaviour, selecting rakes which could only be used later at an alternative location, after an imposed time delay, and foregoing immediate rewards in order to select rakes for future reward retrieval. These findings may not be considered as representing complex prospection given the applicability of learned associations in accounting for much of the current findings, with these accounts discussed in more detail below; however, gibbons are capable of some aspects of future-thinking that are the underpinnings of prospective cognition.

Chapter 5 found gibbons selected an un-baited functional rake at delay periods of up to five minutes before this rake was useable to retrieve a food reward at a second location, suggesting the ability to link the temporally and spatially separate events of rake selection and reward retrieval. Of the 12 subjects, 11 passed the training phase, requiring subjects to select between a functional and non-functional rake, without the relationship between rakes and reward visually present. During an assessment of planned tool use, Dekleva, van den Berg, Spruijt and Sterck (2012) found macaques (*Macaca fascicularis*) had a context specific preference for rakes in the original raking in training context, placed on a testing platform with a food reward. Subjects required progressive step-wise training in order to reliably transport and use functional tools, being rewarded after incremental steps of first manipulating and then gradually transporting tools from a selection area to test compartment. These findings suggest macaques failed to link the distinct events of rake selection and reward retrieval. In contrast, several gibbons displayed a spontaneous understanding of the training task, suggesting these subjects connected the temporally separate events without requiring shaping of their behaviour.

Previous assessments have found gibbons display an increased range of exploratory behaviours in response to novel objects than macaques (*Macaca mulatta*), and perform comparably to chimpanzees (*P. troglodytes*) on problem solving and objects manipulation tasks (Beck; 1974; Bernstein Cunningham et al., 2006; Schusterman & Sharpe, 1963). As suggested by Parker (1974), the increased relative brain size across ape species may result in a more diverse repertoire of available behaviours, allowing for increased learning and problem solving abilities when compared to monkey species. Cunningham et al., (2006) suggested the lack of repeated unproductive manipulations displayed by *Hoolock* gibbons (n = 4) during a raking task may suggest subjects mentally represented the problem before acting, rather than learning through trial and error. The current findings support these previous assessments in suggesting gibbons displayed an increased ability to keep in mind the future reward retrieval event, in comparison to monkeys who required behavioural shaping in order to learn to connect distinct events.

A series of time delays were then introduced between rake selection and reward retrieval, which now occurred at a second platform, assessing if gibbons would continue to link these events once a spatial and increased temporal distance was imposed between them. Ten subjects participated in these experiments, and all reached criterion level performance across 0 seconds, 30 second, and two minute delay periods. The number of trials given with a five minute delay period was limited to ten, with eight of the ten subjects selecting the functional rake significantly above chance levels during these trials. This suggested the increasing delay did not significantly affect subjects' ability to keep in mind the association between rake selection and the future reward retrieval event. Comparing performance during the first ten trials of each experiment, performance levels were slightly lower during the zero delay period trials, suggesting some subjects may have been re-learning the association between the rake and reward. Furthermore, some subject's frequency of functional rake selections decreased as the delay period increased, suggesting a cognitive limit within which these individuals could relate rake selection to reward retrieval. However, as a group, once subjects had learned to associative the temporally and spatially separate events during the zero delay trials, performance levels were consistently high across the initial trials of experiments with a 30 second, two minute and five minute delay period. Likewise, time taken to first productive rake contact was similar across delay periods, suggesting subjects were not required to spend longer considering the task before making selections as this delay increased. Although performance was generally unaffected, motivation to participate in the task may have been affected by the increasing delay, with the time taken to first approach the apparatus increasing as the delay became longer. Delay length significantly predicted the amount of time subjects spent attending to the task during this period, with less time spent engaged with the task as the delay increased. This may also reflect decreasing interest in the task as delay period increased; however, the finding that gibbons did not frequently return to the apparatus during longer delays suggests subjects were not struggling to keep in mind the relation between the rakes and rewards at longer delay periods.

Due to the extended experience with the functional rake over Chapters 3 and 4, is it likely subjects had learned to associate this rake with a reward. This task therefore does not necessarily require a sense of the future, as defined by Raby and Clayton (2009); however, the task is still prospective in nature, as subjects are guiding future actions by learning which behaviours to execute in order to lead to these actions (Suddendorf, Bulley & Miloyan, 2018). Dekleva et al. (2012) suggest associative learning may not fully account for monkeys'(*M. fascicularis*) behaviour during a planned tool use task, as delay periods of up to 20 minutes were imposed before reward retrieval. However, the progressive training and, as with gibbons, increased number of trials, mean associative accounts cannot be ruled out. Monkeys transportation of functional rakes was likely instigated by associatively learned connections to the future reward retrieval event. Given that gibbons may have been guided by these learned associations, their performance is not comparable to that of the large apes, which have been found to select and transport tools over extended delays, and to select novel but functional tools with which they have no prior experience (Mulcahy & Call, 2006; Osvath & Osvath, 2008). Gibbons, like monkeys, displayed basic prospective cognition, connecting temporally and spatially distinct events of rake selection and reward retrieval in order to secure future rewards by learning which actions lead to this outcome. The current findings situate gibbon's prospective abilities closer to those of monkeys, rather than the large apes; however, further assessment may provide evidence of prospection requiring a sense of the future more akin to that seen within large apes.

7.2.3 Self-control

Self-control is an integral component of prospective cognition; even if an individual is capable of foresight and mentally constructing a future event, this may not be acted upon without the ability to inhibit current desires in favour of future ones. Chapter 6 presented an initial investigation into gibbons' self-control abilities. Inter-temporal choice tasks requiring subjects to select between an immediate small or delay large reward, both visibly present, may result in subjects selecting the larger reward due to difficulties in inhibiting selections of larger quantities of items, rather than reflecting an intentional inhibitory response (Beran et al., 2014). The use of rakes, paired with a small reward, aimed to circumvent this issue. In line with previous assessments of self-control in primates (e.g., Anderson, Kuroshima & Fujita, 2010; Beran et al., 2016b; Evans & Beran, 2007), there were individual differences in performance, with half of the subjects failing to inhibit selection of the small reward once this was placed within reach. The remaining three subjects successfully refrained from selecting this reward, instead selecting a rake in order to retrieve a larger, qualitatively different reward, even when this reward was not presented in the same location as the tool, as has been found in large apes (Osvath & Osvath, 2008).

Six of the eight subjects passed training, refraining from selecting an easily accessible small reward in order to use a functional rake to draw in an out of reach large reward, suggesting subjects were able to inhibit selection of a closer reward when the large reward was visible. Experiment 1 presented the functional rake and small reward within reach on platform 1, and an out of reach large reward on platform 2, with the experimenter taking the functional rake to this platform for reward retrieval if selected. All subjects failed this task, potentially due to either a lack of understanding of the task demands or difficulty in displaying self-control. A reinforcement experiment was therefore conducted, aiming to make the relationship between the functional rake and large reward positioned on platform 2 clear to subjects. All subjects reliably selected an out of reach functional rake and large reward, positioned on platform 2, rather than an out of reach small reward on platform 1.

Experiments 2, 3 and 4 followed, aiming to assess if subjects' failure during Experiment 1 had been due to a lack of understanding of the task demands, or alternatively represented a lack of self-control capacities. It was expected the incremental movement of the small reward, closer to the reachable distance of the rake handle, would produce an increasing difficulty in inhibiting selection of this reward, allowing for assessment of at which point subjects failed to display self-control. The experimental set-up during these experiments was similar to Experiment 1, presenting the small reward and functional (non-baited) rake on platform 1, and the large reward on platform 2. Experiment 2 placed the small reward out of reach, before Experiments 3 and 4 incrementally moved this reward back to within reach and closer to the position used during Experiment 1. Three subjects displayed self-control, inhibiting selection of the more immediately available small reward and reliably selecting the functional rake across these experiments. These subjects also reached criterion level when re-presented with Experiment 1, suggesting their initial failure was not due to a lack of self-control abilities. The remaining three subjects failed this task once the small reward was within reach, despite being placed at a distance further from them than the functional rake handle, with these subjects failing to inhibit selection of the more immediate reward. Further testing suggested one of these unsuccessful individuals had difficulty inhibiting the desire for the immediately attainable reward, even when small and large options of a preferred reward were offered. The second subject now displayed a side bias, possibly due to decreased interest in the task, resulting in selection of the functional rake on around 50% of trials. The third reliably selected the functional rake following further testing. It is possible increased experience of the task developed the ability to display self-control in this subject. Alternatively, the presentation of a choice between rewards of different quantities, rather than between rewards differing in both quantity and quality, may have presented an easier task.

The three successful subjects selected the rake significantly across Experiments 2-4 suggesting the closer positioning of the small reward did not affect their ability to inhibit selection of this reward. When compared to the initial presentation, these subjects displayed an improved performance during the repeat of Experiment 1. It is possible the experiences gained through Experiments 2-4, which aimed to present an easier inhibition task than Experiment 1, aided subjects in learning to display self-control during Experiment 1 (Beran et al., 2016b). Alternatively, given that these subjects selected rake at significant levels across all trials in Experiments 2-4, their initial failure may have been due to not understanding the task demands. It is possible these subjects would have succeeded with the experimental set-up of Experiment 1 following the reinforcement experiment. Further research is needed to assess the capabilities of gibbons when presented with two immediately obtainable options.

The presentation of two items at an equivalent distance resulted in a more difficult self-control challenge, with subjects displaying lower performance levels during the repeat of Experiment 1 than they had across Experiments 2-4. It is possible subjects were applying the rule based system of selecting whichever item was closer to them during Experiments 2-4. This is supported by the findings that frequency of large reward selections dropped during the initial trials of Experiment 6, which presented a functional rake at a closer distance than the alternative choice of an out of reach large reward. When presented with two baited rakes, Fujita et al. (2003) found capuchins (*C. apella*) tended to select the rake baited with a reward positioned closer to the subject, regardless of whether this rake was functional for reward retrieval. This bias appears to also be present in gibbons; potentially representing an innate tendency to select closer items, an adaptation which would be beneficial in most natural situations. When presented with two items placed at an equivalent distance, during Experiments 1 and 5, gibbons, like capuchins, did display task sufficient performance and select the rake leading the large reward retrieval, suggesting the ability to make reward directed selections (Fujita et al., 2003).

The performance of the successful subjects suggests self-control is within the cognitive capabilities of the small apes, with the ability to inhibit consumption of more immediate rewards in order to gain access to delayed larger rewards similar to that displayed by large apes and New World monkeys (Dekleva et al., 2012; Osvath & Osvath, 2008). However, the current findings do not provide evidence as to whether subjects mentally envisioned future reward retrieval when making a choice between an immediate small reward now or a large reward accessible by using the rake at an alternate location. Osvath and Osvath (2008) suggest the conflict of current and future desires produced by self-control tasks may induce a mental representation of the future task. However, in order to assess this future thinking, self-control tasks must present rewards of different qualities, and the stimuli related to the future event must not be cued by an associatively learned connection to the future reward. It is possible gibbons' responses during Experiments 1-4 of Chapter 6 were guided by associations with the functional rake, as they had extensive experience with the rake leading to a reward across Chapters 3, 4 and 5. However, the results of Experiment 5 suggested gibbons' selections were reward directed rather than based on a general association between the rake and a reward, with subjects selecting a large reward rather than the functional rake when both these items were placed at an equivalent distance.

Evans and Westergaard (2006) proposed tool use may be related to self-control, with both behaviours requiring foresight and deliberate control over one's actions. In line with this proposal, and Evans and Westergaard's (2006) finding that experienced tool using subjects (*C. apella*) displayed higher levels of self-control than those with less experience, it is possible the experience of using rakes to retrieve rewards over previous chapters aided subjects in displaying self-control. Indeed, the imposed delays between rake selection and reward retrieval during Chapter 5 delayed gratification up to five minutes. This experience of selecting an un-baited functional rake may have aided subjects' performance during Chapter 6, where the additional difficulty of inhibiting selection of an immediately attainable small reward was presented. Overall, the behaviours displayed by gibbons were prospective in the sense that selections were generally not immediately beneficial, but served to retrieve a slightly delayed larger reward. In contrast to Evans and Westergaard's (2006) proposal, it is not suggested here that gibbons acted with foresight, or a sense of future in mind, given that all associative learning accounts cannot be ruled out. Despite this, current findings suggest the three successful gibbons were capable of self-control, a component process of prospective cognition allowing subjects to inhibit selection of immediately attainable rewards in order to select functional rakes to retrieve a large reward.

Delaying gratification may not always be an adaptive behaviour in the wild where animals have to consider predation, social competition, perishable resources, and uncertain future outcomes (Beran & Hopkins, 2018). However, increased inhibitory control has been found to correlate with general intelligence measures, potentially aiding with cognitive processing and intelligent decision making (Beran & Hopkins, 2018). This increased flexible intelligence may allow individuals to recognise when the benefits of self-control outweigh the potential costs. Gibbons may not face the same competitive in-group social pressures as species with larger social groups and dominance hierarchies, given their primarily monogamous, family group structure. Gibbons do, however, show overlap in their wider home range areas, with inter-group encounters and feeding trees shared between groups in these areas, in contrast to the core home range area which is strictly defended (see Cheyne, Capilla, Supiansyah, Cahyaningrum & Smith, 2019). Gibbons also show intra- and inter-specific dietary flexibility, often in response to seasonal variability in available resources (Elder, 2009). Thus, self-control may still be beneficial in wild settings for the small apes, allowing them to pass by less nutritious food sources, or those within their core home range, directing travel paths to preferential or contested feeding patches first. Given the small sample size tested here, and that only half of the subjects successfully refrained from selecting an immediately available reward, further research is needed in order to situate gibbons self-control capacities within the wider primate lineage.

7.3 Genus differences

There was little consistency seen in performance between the tested genera across the presented experiments, with the small numbers progressing to each level making robust evaluation challenging. No genus consistently reached criterion level faster, or selected the functional rake more frequently, than other genera; however, some differences were seen between the experiments of Chapters 3 and 4, where food rewards were more immediately available, and Chapters 5 and 6, which presented delayed rewards.

Hoolock subjects displayed a trend for improved performance across Chapters 3 and 4, with these individuals generally spending less time in the target area than other genera, and being faster to first productive contact. Qualitatively, these subjects appeared to spend much less time considering the task before making selections, with faster, more focussed, rake manipulations made by both *Hoolock* and *Nomascus* subjects than by the other species. These findings support those of Cunningham (2006), who assessed gibbons (*Hoolock*, *Nomascus*, *Symphalangus* and *Hylobates* sp.) capabilities during a raking in task similar to that presented in Chapter 3 and 4. In line with the faster behaviour displayed by *Hoolock* subjects in the current research, Cunningham (2006) found *Hoolock* subjects were fastest to first solution. However, in contrast to the current findings, where *Hoolock* subjects generally spent less time in the target area, Cunningham (2006) also found *Hoolock* subjects were the most attentive to the task, rarely leaving the target area. Given the similarity of the tasks, it is possible that the increased age of *Hoolock* subjects in the current research led to less attentive behaviour. The average age of *Hoolock* subjects assessed by Cunningham (2006) was 5.8 years, in comparison to an average age of 11.4 years during pilot testing (Chapter 3). In line with previous reviews suggesting adult primates may display increased problem solving abilities (Griffin & Guez, 204; Reader & Laland, 2001), it is possible the increased age of the current subjects reflected increased object manipulation skills, resulting in a shorter learning period required in order for these individuals to attend to the relevant features of the rakes (see section 7.4 for further discussion of age group differences in performance).

During the experiments of Chapter 3 and 4, *Hylobates* subjects, anecdotally, appeared to be less engaged with the task, with these subjects frequently 'playing' with the rakes and displaying less direct manipulations. Given the low number of *Symphalangus* subjects tested (two in Chapter 3 and one in Chapter 4), it is difficult to draw conclusions from these subjects behaviour; however, siamangs generally displayed more abrupt behaviour with the rakes during pilot testing, displaying little consideration of the task before selection. Marlow, the sole siamang included following pilot testing displayed decreasing levels of interest in the task. Speculatively, it is suggested this was due to the death of her mother, who she was housed with, suggesting her performance levels are not representative of this species abilities.

Differences in ecology may account for the different approaches to the tasks displayed by *Hoolock*, *Nomascus* and *Hylobates* species. *H. leuconedys* are found in more northern locations, such as Myanmar, India and China, with *N. leucogenys* found in Laos, Vietnam and China (Brandon-Jones et al., 2004; Chetry, Chetry, Ghosh & Singh, 2010; Harding, 2012). In contrast, *H. pileatus* are found in more southern locations of Cambodia, Laos and Thailand, with *H. moloch* endemic to the island of Java (Brandon-Jones et al., 2004). Fan, Ai, Fei, Zhang and Yuan (2013) report *H. leuconedys* living in a northern montane forest displayed an increased flexibility in diet when compared to gibbon species found in the tropics. These individuals adopted energy-conserving strategies in response to seasonal fluctuations in temperature and food availability, eating increased amounts of insects and leaves when fruit became scarce. In contrast, lower seasonal fluctuations in diet have been found for *Hylobates* species, with fruit comprising their main food source across most months (Bartlett, 2009; Kim, Lappan & Choe, 2012). The difference in seasonal variation in both food availability and weather conditions between more northern and southern species may therefore result in different levels of attentiveness to resources and investigative behaviour. This increased attentiveness may have led to the generally better performance displayed by *Hoolock* and *Nomascus* subjects across Chapters 3 and 4, learning which features of the rakes to attend to faster. In contrast, the decreasing requirement for receptiveness to potential resources may have resulted in the generally lower performance levels of

Hylobates subjects, requiring a longer learning period in order to attend to the relevant perceptual features.

In contrast to the trends found in earlier chapters, *Hylobates* subjects performed better than both *Hoolock* and *Nomascus* subjects across Chapters 5 and 6. *Hylobates* subjects also generally took longer to first productive rake contact during the time delay experiments of Chapter 5, and spent less time attending to the task during the delay periods, potentially signalling less impatient behaviour in this genus. It is possible the discussed differences in ecology contributed to *Hylobates* subjects tolerating the delay in reward retrieval across these experiments better than the other genera. This is supported by the finding that two of the three individuals who successfully completed all self-control experiments during Chapter 6 were also *Hylobates* subjects, again suggesting higher levels of inhibitory control in this genus. That the increased responsiveness to resources found in more northern located species may contribute to decreasing levels of self-control is supported by the findings that two of the individuals who failed to refrain from selecting the small reward once it was within reach during self-control tasks (Experiment 3, Chapter 6) were *Hoolock* subjects. Cunningham (2006) found *Hylobates* subjects were slowest to first solution during a raking in task, with these subjects described as timid and lacking motivation, a suggestion supported by the current findings that these subjects generally appeared to less engaged with the task than other genera. Although this behaviour may result in poorer behaviour during tasks where more immediate responses are possible, the slower, more patient behaviour appeared to aid *Hylobates* subjects during tasks requiring tolerance of delays, suggesting increased levels of impulse control in this genus.

7.4 Age and gender differences

Across Chapters 3, 4 and 5, sub-adults generally spent less time engaged with the apparatus and displayed lower frequencies of functional rake selections than adults. During pilot testing (Chapter 3) age was classified as adult (8+ years), sub-adult (6-8 years) and adolescent (3-5 years) (as classified in Palombit, 1994). Age was classified as either sub-adult (<8 years) or adult (>8 years) for the remaining chapters.

Adolescents spent the most time attending to the task during pilot testing, supporting previous findings that younger *Callitrichid* monkeys displayed increased attentiveness to novel objects (Kendal, Coe & Laland, 2005). However, in contrast to Kendal et al.'s (2005) findings, across Chapters 3, 4 and 5 adults tended to spend more time engaged with the task than sub-adults and were slower to first productive rake contact. Differences in task type potentially led to these findings, with the extractive foraging task presented by Kendal et al. (2005) requiring more natural behaviour than the current task. The increased foraging experience of adult subjects, given their life history, may have resulted in the decreased attentiveness to the task and faster successful manipulations found in adult monkeys (Kendal et al., 2005). In contrast, there are no reports of wild gibbons using objects to retrieve out of reach food sources (Shumaker et al., 2011) with this less natural behaviour potentially more difficult for gibbons, resulting in the decreasing levels of interest shown by sub-adults. Furthermore, in contrast to the foraging task, during Chapter 5 the reward was absent during the rake selection period. This task required subjects to keep in mind the absent reward during rake selections, and it is possible sub-adults displayed lower levels of competency at this than adults, supported by their generally worse performance on the task, resulting in decreased interest in the un-baited apparatus.

Levels of inhibitory control may also have contributed to the generally faster time to contact the rakes and decreased time spent engaged with the apparatus shown by younger subjects. In humans, the ability to delay gratification develops throughout childhood and into adolescence (e.g., see Mischel & Metzner, 1962). The ages of three to six years are thought to be important transitional periods for the development of inhibitory control, in relation to the maturation of brain regions involved in self-control (Tarullo, Obradović & Gunnar, 2009). Inhibitory control in non-human large apes has been found to be comparable to that of children between the ages of three and five years, with uniquely human levels of self-control proposed to develop around the age of six years (Herrmann, Misch, Hernandez-Lloreda & Tomasello, 2015; Vlamings, Hare & Call, 2010). There are few assessments of the development of self-control in gibbons, but it is possible this ability follows the similar developmental trajectories as with large apes, with the maturation of brain regions throughout infancy and adolescence leading to improved inhibitory control. This may have contributed to the decreased time spent considering the task before making selections displayed by younger subjects. In contrast, older individuals may have been capable of spending longer processing the task demands before making selections as a result of increased self-control abilities.

Time to first approach the apparatus was also measured during Chapter 5. Sub-adults were marginally slower than adults to first approach during the training task. As before, it is possible the lack of a visible reward during this task initially decreased interest in sub-adult subjects. This was also the only experiment where sub-adults generally selected the functional rake more frequently than adults. Adults may have learned to attend to the rakes in combination with the rewards during Chapter 3 and 4, whereas sub-adults attended to the more simple perceptual features of the rakes, resulting in adult subjects re-learning which aspects of the rake to attend to without the presence of the reward. The finding that sub-adults were marginally faster than adults to approach the task during the experiments with imposed delay periods suggests these subjects displayed increased interest in the task once they had experienced reward retrieval. This again may reflect a decrease in inhibitory control by sub-adults or, alternatively, it is possible this reflected an increased general interest in the apparatus when it was first presented, in line with Kendal et al.'s (2005) findings. As time to first approach was not measured across Chapters 3, 4 or 6, it is not possible to say whether this was a consistent trend.

The generally improved performance of adult subjects across the current research contrasts with Cheyne (2009), who found that of the six gibbons (*Hylobates* sp.) who successfully retrieved a food reward from a puzzle feeder, the younger individuals spent longer interacting with the task than the adult subjects. Sub-adults also performed better on the puzzle task than adults, retrieving a greater number of rewards (mean = 1.25/4) than adults (mean = 0.8/4). As with the findings of Kendal et al. (2005), differences in task type may have led to the differences in performance levels and attentiveness between the current research and Cheyne (2009). This suggestion is supported by the findings of Cunningham (2006) presenting gibbons (*Hoolock*, *Nomascus*, *Symphalangus* and *Hylobates* sp.) with a similar raking in task to that used during Chapters 3 and 4. In line with Cheyne's (2009) findings, Cunningham (2006) found time spent in the target area declined as age increased. However, despite this increased time spent in the target area, Cunningham (2006) found younger subjects (<10 years) were not more likely to reach solution than older (>10 years) individuals, suggesting increased engagement did not improve performance in younger individuals.

Assessments of gibbon's abilities during object permanence tasks, and learning of colour-reward associations have also found no significant difference in performance between age groups (D'Agostino & Cunningham, 2015; Fedor, Skollár, Szerencsy & Ujhelyi, 2008). In line with these assessments, the difference in performance levels between age groups in the current research was only significant during Experiment 1 of Chapter 4. However, the general trend of adults displaying a superior performance to sub-adults supports research suggesting increased problem solving and innovative behaviours in adults, potentially as a result of increased experience and competence in older individuals (Griffin & Guez, 2014; Reader & Laland, 2001). In contrast to Cunningham (2006), adults in the current research spent longer engaged with the task than younger individuals, it is possible this increased time spent considering the task before making selections contributed to the slightly improved performance of adults.

Across Chapters 3, 4 and 5, females displayed a consistent trend for selecting the functional rake at higher levels than males. Females were also slower to first productive rake contact and generally spent longer in the target area attending to the apparatus during Chapters 3 and 4 and the delay experiments of Chapter 5. These findings support those of Cheyne (2009), who found females (*Hylobates* sp.) retrieved more rewards (mean = 1.33/4) than males (mean = 0.83/4), and that of the successful subjects, males spent less time interacting with the task than females. These findings also partially support those of Cunningham et al. (2011) who found gibbons (*Hoolock*, *Nomascus*, *Symphalangus* and *Hylobates* sp.) displayed a sex difference in the effect of prior apparatus experience on time to solution during a raking in task. Cunningham et al. (2011) found females with no prior experience were slower to task solution than either males with no prior exposure, or both males and females with prior experience. Time to solution was not measured in the current research; however, Cunningham et al.'s (2011) suggestion that females may be more cautious of novel objects is supported by current findings, with females generally spending more time in the target area but being slower to first productive rake contact.

That females may be more cautious when faced with novel situations is also supported by work addressing innovation in primates, with Reader and Laland (2001) finding females displayed lower levels of innovation than males. Given that innovative behaviour may be risky, requiring exploration of novel objects or locations, this may represent more risk-aversive behaviour in females. Females may have more to lose, in fitness terms, by exploring novel objects or locations. In contrast to the current research, Cunningham et al. (2011) found that males were slower to first contact the apparatus. Likewise, Cheyne (2009) found males were slower to approach the puzzle task than females. It is possible that the relatively high number of females with dependent offspring in the current research, a number of whom were tested in their family groups, influenced the levels of caution displayed by these female individuals. Having dependent offspring has been found to alter the behaviour of females during group movement in the wild. Carpenter (1940) noted that females (*H. lar*) carrying offspring were usually found at the rear of the group during group locomotion, whereas those without dependent offspring are frequently positioned at the front. This may be due to decreased mobility as a result of carrying an infant, but may also reflect more cautious behaviour on behalf of the female.

Sex differences have also been reported to influence travel patterns and leadership during group movements within gibbons. During a long term study with lar gibbons (*H. lar*) Reichard and Sommer (1997) found groups are structured by female resource-defence, and male mate-defence. Females were found to lead the majority of travels into and out of feeding trees, whereas males led the group during inter-group encounters. If females are primarily responsible for locating food, it may be expected that males would have benefitted more from witnessing females complete testing first during the tasks presented throughout this thesis, requiring object manipulation in order to obtain a food reward. However, considering the effects of testing order, in enclosures where one male and one female were tested and there was a significant difference in performance levels, males did not consistently perform better after witnessing a female taking part in testing. Males who were tested second performed better than their female partner on only 37.5% of such cases, whereas females tested second performed better on 50% of cases. This may be reflective of increased sensitivity to potential resources in females, whereas males may be predominantly focussed on territorial defence. The increased responsiveness to resources, in combination with the increased time spent considering the task before making selections, may explain the generally improved performance seen by female subjects throughout the current research.

7.5 Future research

The current thesis presented preliminary investigations into basic prospective abilities in gibbons; namely the selection of functional objects for future use and the capacity to display self-control. With regards to investigations of self-control, it would be of interest to assess these capacities in gibbons with less object manipulation experience than the current subjects. In line with Beran et al.'s (2016b) finding that capuchins (*C. apella*) with previous cognitive testing experience displayed improved inhibitory control over repeated testing in self-control tasks, it is possible the experience of selecting un-baited functional rakes and tolerating delay periods before reward retrieval during Chapter 5 aided subjects in Chapter 6. Although some task experience is required in order for subjects to understand the task demands, comparing these abilities between highly and less experienced subjects would allow for investigation of whether the current levels of self-control displayed by gibbons was related to their object manipulation experience or, alternatively, whether this component of prospective cognition is present within task naïve gibbons.

Previous assessments of planned tool use have required subjects to transport and properly orient tools before using them, as in true tool use (see Shumaker et al., 2011) Cunningham (2006) found little evidence for true-tool use capabilities in gibbons (*Hoolock*, *Nomascus*, *Symphalangus* and *Hylobates* sp.) when required to re-orient tools in order to draw in food rewards. However, Dekleva et al. (2012) found several monkeys (*M. fascicularis*) transported and correctly oriented tools during a raking in task following step-wise training. It is possible that given progressive training, gibbons could also succeed at this task. Assessment of these abilities would allow for a clearer comparison of planned tool use between gibbons and monkeys (Dekleva et al., 2012) and large apes (Mulcahy & Call, 2006; Osvath & Osvath, 2008).

The current findings of Chapters 5 and 6 do not suggest gibbons mentally represented future reward retrieval, given associative accounts of behaviour cannot be ruled out (Raby & Clayton, 2009; Suddendorf & Corballis, 2010). Gibbons' selections of the functional rake for future reward retrieval during Chapter 5 may have been due to learned stimulus-response relations between the functional rake and future reward, given the extensive experience of this rake reliably leading to a reward across previous experiments. The use of novel functional objects, which subjects did not have extensive previous experience with, would assess whether gibbons can mentally represent a future reward retrieval event without the presence of a learned association. This suggestion is in line with Osvath and Osvath (2008), who found two chimpanzees (*P. troglodytes*) and an orangutan (*P. abelii*) successfully selected a novel functional tool when presented with the novel tool, two novel distractor items and one familiar item previously used for food retrieval but non-functional with the current apparatus. The two novel distractor items differed markedly in appearance from the other items; for example, a small teddy bear and transparent ruler were paired with a multi-coloured rubber hose (functional tool) and a non-hollow bamboo stick (familiar distractor object). Given gibbons ability to attend to functionally relevant features of rakes with marked perceptual differences (Chapter 4), it is possible they would also attend to the functionally relevant features of novel items and select an appropriate item for future use.

During the experiments of Chapter 6, the large reward remained visible during the selection period, likely inducing a desire for this reward and cueing subjects' selections of the rake. Success at this task without the visual presence of the large reward would provide stronger support for the ability to keep in mind future reward retrieval. These suggestions would build on the future-oriented abilities found within gibbons in the current research, assessing if these apes possess higher levels of prospective cognition more akin to the ability to mentally represent future events displayed by large apes.

7.6 Conclusions

This thesis presents an initial investigation into components of prospective cognition within the small apes, with support for basic prospection and some evidence of self-control providing preliminary support for evolutionary continuity in these abilities across apes. Gibbons hold a unique phylogenetic position, intermediate between the large apes and monkeys, and as such present an important opportunity to assess the evolutionary emergence of cognitive capacities within primates. Given that gibbons are the most species-rich group of apes, they allow for diverse cross-species, and cross-genera, comparisons. Despite this, gibbons are a vastly understudied group of primates, with numerous reports of low motivational levels in these apes. For example, Harlow, Uehling and Maslow (1932) found a gibbon (*H. lar*) displayed low motivation levels during a delayed response task. This subject was described as disinterested in the task overall and reluctant to respond when delay periods of a minute or more were introduced between apparatus baiting and the selection period. Berkson (1962) suggested low motivational levels in gibbons may be counteracted by regulating subjects' appetite and providing high quality rewards. Subjects were not food deprived during the current experiments, being fed multiple times throughout the day, often recently before testing. Despite this, gibbons were motivated to complete testing across all experiments. The findings of Chapter 5 in particular contest reports of low motivation in these apes, with subjects enduring delays of up to five minutes for a relatively small reward of three blueberries. Gibbons were slower to enter the target area as the delay period increased, suggesting decreasing motivation with the longer time delays; however, subjects continued to reliably participate in the experiments. The current findings suggest with patience and methodology that allows gibbons to understand the task demands, they can be motivated to participate in cognitive testing. Therefore, inclusion of gibbons as test subjects in comparative psychology is encouraged.

Spaulding and Hauser (2005) proposed non tool-using primates may possess an innate mechanism for recognising the functionally relevant features of objects, with task relevant experience required in order for the mechanism to function effectively. The current findings support this proposal as, in line with findings from other typically non-tool using species (Hauser, 1997; Hauser et al., 2002; Spaulding & Hauser, 2005), gibbons were able to attend to the functionally relevant features of rakes following experimental experience. A number of subjects were able to differentiate between functional and non-functional rakes, and could take into account the three dimensional relationships between rakes and food rewards, performing comparably to chimpanzees (*P. troglodytes*) on this task. However, there needed to be a period of learning before performance reached criterion level, suggesting gibbons required experience of experiments where the rewards position relative to the rake could be taken in to account in order to succeed during a partial replication of the inverted rake task (Povinelli & Reaux, 2000). Gibbons do not routinely use objects to obtain out of reach food sources in the wild. Although there are some reports of pulling in vines or branches to reach food, these apes do not display behaviours such as termite dipping found within the large apes (Shumaker et al., 2011). Furthermore, gibbon's social structures do not provide the opportunity for extensive social transmission of skills seen within species which live in larger communities. Thus although the current findings suggest gibbons have the ability to attend to relevant features of objects, their natural habitat and sociality may not present the need for more complex object manipulation or tool use skills to develop in wild populations.

Once gibbons had learned the necessary skills to select a functional rake to obtain a food reward, a series of experiments investigated their capacity for prospective cognition. Suddendorf and Corballis (1997; 2010) suggest mental time travel is a uniquely human trait. Debate as to whether non-human primates share this ability has diverted attention from assessment of the range of prospective cognition capabilities present within various species (Osvath & Martin-Ordas, 2014; Raby & Clayton, 2009). The current research presented initial investigations of prospective cognition levels within the small apes. Gibbons displayed basic prospective cognition, relating the temporally and spatially distinct events of rake selection and reward retrieval. Performance was not comparable to assessments of planned tool use in large apes, given that stimulus-response associations cannot be ruled out. However, gibbons appeared to link the distinct events of rake selection and reward retrieval without progressive training, in contrast to monkeys (*M. fascicularis*) (Dekleva et al., 2012).

Three gibbons displayed self-control, foregoing an immediate small reward in order to select a functional rake useable for future retrieval of a larger reward. Gibbons displayed a tendency to select the closest item, a bias also found in capuchins (*C. apella*) (Fujita et al., 2003); however, subjects behaviour appeared to be reward directed, selecting the large reward both when this did, and did not, require selection of the functional rake. Subject's selections were likely cued by the presence of the large reward during the self-control tasks, as a result, their behaviour cannot be considered as requiring a sense of the future. This contrasts with the performance of large apes, which have been found to inhibit selection of a small reward in favour of a functional tool they were required to transport and retain over extended time delays in order to retrieve a future reward (Osvath & Osvath, 2008).

Restrictive criteria such as that set out by Suddendorf and Corballis (2007; 2010) may result in a narrow approach to prospective cognition. This research considered more basic prospection, and assessed some of the individual component processes inherent in future directed behaviour. The current findings suggest basic prospection is within gibbons' capabilities, with some proficiency also found with self-control, an important pre-requisite to much future oriented behaviour. Given the potential for learned associations accounting for subjects' behaviour, these findings do not provide unequivocal evidence for gibbons possessing a level of prospection requiring a sense of the future, comparable to the more complex prospective cognition displayed by large apes (Mulcahy & Call, 2006; Osvath & Osvath, 2008). However, the current findings do not provide evidence for an absence of these abilities either. Self-control and future directed thinking may be of benefit to gibbons in natural settings, where individuals are required to keep track of territory boundaries, neighbouring groups, resource locations and availability. The support for gibbons possessing cognitive maps of their territory, increasing their foraging efficiency by planning movements between preferred food sources suggests gibbons do display future oriented behaviour in the wild (Asensio et al., 2011). However, this again does not necessarily suggest these individuals act with a sense of the future, with, for example, natural landmarks associated with specific food sources perhaps used to guide each sequential movement rather than a mental representation of future travel routes. Given the support for basic prospection in gibbons, further research incorporating behavioural criteria such as that suggested by Osvath and Osvath (2008) is required in order to assess higher levels of prospection, providing a clearer comparison of abilities between large and small apes.

Appendices

Appendix 1: Ethics committee approval letter

JM/CW/CR/SHS/14/P/011

19th November 2014

[redacted]

Dear Heather

An Investigation of Flexible Prospective Cognition in Gibbons

This is to notify you that the Ethics Committee have looked at your submission and you have been granted **full ethical approval** to collect data for your project as entitled above. This is subject to the following standard conditions:

i You must remain in regular contact with your project supervisor

ii Your supervisor must see a copy of all experimental materials and your procedure prior to commencing data collection

iii If you make any substantive changes to your project plan, you must submit a new ethical approval application to the Committee. Application forms and the accompanying explanatory document are on the Intranet. Completed forms should be handed in to the School Office, School of Social & Health Sciences, Level 5, Kydd Building, Dundee.

iv Any changes to the procedures must be negotiated with your supervisor

The Committee also made the following comments:

Ethical approval is recommended subject to adherence of observational research procedures at the Gibbon Conservation Center.

There are separate university forms for fieldwork/travel that you would need to fill in for insurance purposes before you go. (These are all available from Claire McCurley in Estates if you are unsure where to find them on portal).

Failure to comply with these conditions will result in your ethical approval being revoked by the Ethics Committee.

Should you have any queries please contact your Supervisor.

Yours sincerely

School Ethics Committee
School of Social & Health Sciences

References

Andayani, N., Morales, J. C., Forstner, M. R. J., Supriatna, J. & Melnick, D. J. 2001. Genetic variability in mtDNA of the silvery gibbon: Implications for the conservation of a critically endangered species. *Conservation Biology, 15*, 770-775.

Anderson, M. R. 2012. Comprehension of object permanence and single transposition in gibbons. *Behaviour, 149*, 441-459.

Anderson, J. R., Kuroshima, H. & Fujita, K. 2010. Delay of gratification in capuchin monkeys (*Cebus apella*) and squirrel monkeys (*Saimiri sciureus*). *Journal of Comparative Psychology, 124*, 205-210.

Asensio, N., Brockelman, W. Y., Malaivijitnond, S. & Reichard, U. H. 2011. Gibbon travel paths are goal oriented. *Animal Cognition, 14*, 395-405.

Atance, C. M. & O'Neill, D. K. 2005. The emergence of episodic future thinking in humans. *Learning and Motivation, 36*, 126-144.

Barelli, C., Matsudaira, K., Wolf, T., Roos, C., Heistermann, M., Hodges, K., Ishida, T., Malaivijitnond, S. & Reichard, U. H. 2013. Extra-pair paternity confirmed in wild white-handed gibbons. *American Journal of Primatology, 75*, 1185-1195.

Bartlett, T. Q. 2009. Seasonal home range use and defendability in white-handed gibbons (*Hylobates lar*) in Khao Yai national park, Thailand. In S. Lappan & D. J. Whittaker (Eds.), *The gibbons: New perspectives on small ape socioecology and population biology* (pp. 265-275). Springer: New York.

Bean, A. 1999. Ecology of sex differences in great ape foraging. In P. C. Lee (Ed.), *Comparative Primate Socioecology* (pp. 339-362). Cambridge University Press: Cambridge.

Beck, B. B. 1967. A study of problem solving by gibbons. *Behaviour, 28*, 95-109.

Beck, B. B. 1980. *Animal Tool Behaviour: The Use and Manufacture of Tools by Animals.* Garland STPM Press: New York.

Beran, M. J. 2002. Maintenance of self-imposed delay of gratification by four chimpanzees (*Pan troglodytes*) and an orangutan (*Pongo pygmaeus*). *The Journal of General Psychology, 129*, 49-66.

Beran, M. J. & Evans, T. A. 2006. Maintenance of delay of gratification by four chimpanzees (*Pan troglodytes*): The effects of delayed reward visibility, experimenter presences, and extended delay intervals. *Behavioural Processes, 73*, 315-324.

Beran, M. J., Evans, T. A., Paglieri, F., McIntyre, J. M., Addessi, E. & Hopkins, W. D. 2014. Chimpanzees (*Pan troglodytes*) can wait, when they choose to: A study with the hybrid delay task. *Animal Cognition, 17*, 197-205.

Beran, M. J. & Hopkins, W. D. 2018. Self-control in chimpanzees relates to general intelligence. *Current Biology, 28*, 574-579.

Beran, M. J., Menzel, C. R., Parrish, A. E., Perdue, B. M., Sayers, K., Smith, J. D. & Washburn, D. A. 2016a. Primate cognition: Attention, episodic memory, prospective memory, self-control, and metacognition as examples of cognitive control in nonhuman primates. *Wiley Interdisciplinary Reviews: Cognitive Science, 7*, 294-316.

Beran, M. J., Perdue, B. M., Bramlett, J. L., Menzel, C. R. & Evans, T. A. 2012. Prospective memory in a language-trained chimpanzee (*Pan troglodytes*). *Learning and Motivation, 43*, 192-199.

Beran, M. J., Perdue, B. M., Rossettie, M. S., James, B. T., Whitham, W., Walker, B., Futch, S. E. & Parrish, A. E. 2016b. Self-control assessments of capuchin monkeys with the rotating tray task and the accumulation task. *Behavioural Processes, 129*, 68-79.

Berkson, G. 1962. Food motivation and delayed response in gibbons. *Journal of Comparative and Physiological Psychology, 55*, 1040-1043.

Bernstein, I. S., Schusterman, R. J. & Sharpe, L. G. 1963. A comparison of rhesus monkey and gibbon responses to unfamiliar situations. *Journal of Comparative and Physiological Psychology, 56*, 914-916.

Boddy, A. M., McGowen, M. R., Sherwood, C. C., Grossman, L. I., Goodman, M. & Wildman, D. E. 2012. Comparative analysis of encephalization in mammals reveals relaxed constrains on anthropoid primate and cetacean brain scaling. *Journal of Evolutionary Biology, 25*, 981-994.

Brandon-Jones, D., Eudey, A. A., Geissmann, T., Groves, C. P., Melnick, D. J., Morales, J. C., Shekelle, M. & Stewart, C. B. 2004. Asian primate classification. *International Journal of Primatology, 25*, 97-164.

Bräuer, J. & Call, J. 2014. Apes produce tools for future use. *American Journal of Primatology, 77*, 254-263.

Brockelman, W. Y. 2009. Ecology and the social system of gibbons. In S. Lappan & D. J. Whittaker (Eds.), *The gibbons: New perspectives on small ape socioecology and population biology* (pp. 211-239). Springer: New York.

Brockelman, W. Y., Reichard, U., Treesucon, U. & Raemaekers, J. J. 1998. Dispersal, pair formation and social structure in gibbons (*Hylobates lar*). *Behavioral Ecology and Sociobiology, 42*, 329-339.

Buckner, R. L. 2013. The cerebellum and cognitive function: 25 years of insight from anatomy and neuroimaging. *Neuron, 80*, 807-815.

Buckner, R. L. & Carroll, D. C. 2007. Self-projection and the brain. *Trends in Cognitive Sciences, 11*, 49-57.

Burns, B. L. & Judge, D. S. 2016. The varied path to adulthood: Plasticity in developmental timing in hylobatids. *American Journal of Primatology, 78*, 610-625.

Carere, C. & Locurto, C. 2011. Interaction between animal personality and animal cognition. *Current Zoology, 57*, 491-498.

Carpenter, C. R. 1940. A field study in Siam of the behaviour and social relations of the gibbon (*Hylobates lar*). *Comparative Psychology Monographs, 16*, 1-206.

Chan, B. P. L., Mak, C. F., Yang, J. & Huang, X. 2017. Population distribution, vocalisation and conservation of the Gaoligong hoolock gibbon (*Hoolock tianxing*) in the Tengchong section of the Gaoligongshan national nature reserve, China. *Primate conservation, 31*, 1-7.

Chatterjee, H. J. 2006. Phylogeny and biogeography of gibbons: A dispersal-vicariance analysis. *International Journal of Primatology, 27*, 699-712.

Chatterjee, H. J. 2009. Evolutionary relationships among the gibbons: A biogeographic perspective. In S. Lappan & D. J. Whittaker (Eds.), *The gibbons: New perspectives on small ape socioecology and population biology* (pp. 13-36). Springer: New York.

Chetry, D., Chetry, R., Ghosh, K. & Singh, A. K. 2010. Status and distribution of the eastern hoolock gibbon (*Hoolock leuconedys*) in Mehao wildlife sanctuary, Arunachal Pradesh, India. *Primate Conservation, 25*, 87-94.

Cheyne, S. M. 2009. Studying social development and cognitive abilities in gibbons (*Hylobates* SPP): methods and applications. In E. Potocki & J. Krasinski (Eds.), *Primatology: Theories, Methods and Research* (pp. 129-152). Nova Science Publishers: New York.

Cheyne, S. M., Capilla, B. R., Supiansyah, A. K., Cahyaningrum, A. E. & Smith, D. E. 2019. Home range variation and site fidelity or Bornean southern gibbons (*Hylobates albibarbis*). *PLoS ONE, 14*, e0217784.

Chivers, D. J. 1984. Feeding and ranging in gibbons: A summary. In H. Preuschoft, D. J. Chivers, W. Y. Brockelman & N. Creel (Eds.). *The lesser apes: Evolutionary and Behavioural Biology* (pp. 267-281). Edinburgh University Press: Edinburgh.

Choudhury, A. 2013. Description of a new subspecies of hoolock gibbon *Hoolock hoolock* from North East India. *Newsletter & Journal of The Rhino Foundation for Nature in NE India, 9*, 49-59.

Christel, M. 1993. Grasping techniques and hand preferences in Hominoidea. In H. Preuschoft & D. J. Chivers (Eds.), *Hands of Primates* (pp. 91-108). Springer: Vienna.

Clayton, N. S., Correia, S. P. C., Raby, C. R., Alexis, D. M., Emery, N. J. & Dickinson, A. 2008. Response to Suddendorf & Corballis (2008): In defence of animal foresight. *Animal Behaviour, 76*, 9-11.

Clayton, N. S. & Wilkins, C. 2017. Seven Myths of Memory. *Behavioural Processes, 152*, 3-9.

Coleman, K., Tully, L. A. & McMillan, J. L. 2005. Temperament correlates with training success in adult rhesus macaques. *American Journal of Primatology, 65*, 63-71.

Corballis, M. C. 2014. Mental time travel: How the mind escapes from the present. *Cosmology, 18*, 139-145.

Cummins-Sebree, S. E. & Fragaszy, D. M. 2005. Choosing and using tools: Capuchins (*Cebus apella*) use a different metric than tamarins (*Saguinus oedipus*). *Journal of Comparative Psychology, 119*, 210-219.

Cunningham, C. L. 2006. *Cognitive flexibility in gibbons (Hylobatidae): Object manipulation and tool use* (Ph.D. diss.). University of Stirling: Stirling.

Cunningham, C. L., Anderson, J. R. & Mootnick, A. R. 2006. Object manipulation to obtain a food reward in hoolock gibbons, *Bunopithecus hoolock*. *Animal Behaviour, 71*, 621-629.

Cunningham, C., Anderson, J. & Mootnick, A. 2011. A sex difference in effect of prior experience on object-mediated problem-solving in gibbons. *Animal Cognition, 14*, 599-605.

D'Agostino, J. & Cunningham, C. 2015. Preliminary investigation of flexibility in learning color-reward associations in gibbons (*Hylobatidae*). *American Journal of Primatology, 77*, 854-868.

Dallman, R. & Geissmann, T. 2009. Individual and geographical variability in the songs of wild silvery gibbons (*Hylobates moloch*) on Java, Indonesia. In S. Lappan & D. J. Whittaker (Eds.), *The gibbons: New perspectives on small ape socioecology and population biology* (pp. 91-110). Springer: New York.

Damerius, L. A., Burkart, J. M., van Noordwijk, M. A., Haun, D. B. M., Kosonen, Z. K., Galdikas, B. M. F., Saraswati, Y., Kurniawan, D. & van Schaik, C.

P. 2019. General cognitive abilities in orangutans (*Pongo abelii* and *Pongo pygmaeus*). *Intelligence, 74*, 3-11.

Deaner, R. O., Isler, K., Burkart, J. & van Schaik, C. 2007. Overall brain size, and not encephalization quotient, best predicts cognitive ability across non-human primates. *Brain, Behavior and Evolution, 70*, 115-124.

Deaner, R. O., van Schaik, C. P. & Johnson, V. 2006. Do some taxa have better domain-general cognition than others? A meta-analysis of nonhuman primate studies. *Evolutionary Psychology, 4*, 149-196.

Dekleva, M., van den Berg, L., Spruijt, B. M. & Sterck, E. H. M. 2012. Take it or leave it: Transport of tools for future use by long-tailed macaques (*Macaca fascicularis*). *Behavioural Processes, 90*, 392-401.

Dufour, V. & Sterck, E. H. M. 2008. Chimpanzees fail to plan in an exchange task but succeed in a tool-using procedure. *Behavioural Processes, 79*, 19-27.

Elder, A. A. 2009. Hylobatid diets revisited: The importance of body mass, fruit availability, and interspecific competition. In S. Lappan & D. J. Whittaker (Eds.), *The gibbons: New perspectives on small ape socioecology and population biology* (pp. 133-159). Springer: New York.

Evans, T. A. & Beran, M. J. 2007. Delay of gratification and delay maintenance by rhesus macaques (*Macaca mulatta*). *The Journal of General Psychology, 134*, 199-216.

Evans, T. A. & Beran, M. J. 2012. Monkeys exhibit prospective memory in a computerized task. *Cognition, 125*, 131-140.

Evans, T. A., Perdue, B. & Beran, M. J. 2014. The relationship between event-based prospective memory and ongoing task performance in chimpanzees (*Pan troglodytes*). *PLoS ONE, 9*, e112015.

Evans, T. A. & Westergaard, G. C. 2006. Self-control and tool use in tufted capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology, 120*, 163-166.

Fan, PF., Ai, HS., Fei, HL., Zhang, D. & Yuan, SD. 2013. Seasonal variation of diet and time budget of eastern hoolock gibbons (*Hoolock leuconedys*) living in a northern montane forest. *Primates, 54*, 137-146.

Fan, PF., He, K., Chen, X., Ortiz, A., Zhang, B., Zhao, C., Li, YQ., Zhang, HB., Kimock, C., Wang, WZ., Groves, C., Turvey, S. T., Roos, C., Helgen, K. M. & Jiang, XL. 2016. Description of a new species of *Hoolock* gibbon (Primates: Hylobatidae) based on integrative taxonomy. *American Journal of Primatology, 79*, 1-15.

Fedor, A., Skollár, G., Szerencsy, N. & Ujhelyi, M. 2008. Object permanence tests on gibbons (*Hylobatidae*). *Journal of Comparative Psychology, 122*, 403-417.

Fragaszy, D. M., Biro, D., Eshchar, Y., Humle, T., Izar, P., Resende, B. & Visalbergi, E. 2013. The fourth dimension of tool use: Temporally enduring artefacts aid primates learning to use tool. *Philosophical Transactions of the Royal Society of London B, 368*, 20120410.

Fragaszy, D. M., Visalberghi, E. & Fedigan, L. M. 2004. *The Complete Capuchin: The Biology of the Genus Cebus*. Cambridge University Press: Cambridge.

Frisch, J. E. 1963. Sex-differences in the canines of the gibbon (*Hylobates lar*). *Primates, 4*, 1-10.

Fujita, K., Kuroshima, H. & Asai, S. 2003. How do tufted capuchin monkeys (*Cebus apella*) understand causality involved in tool use? *Journal of Experimental Psychology: Animal Behavior Processes, 29*, 233-242.

Garson, G. D. 2013. *Generalized linear models/Generalized estimating equations*. Statistical Publishing Associates: North Carolina.

Geissmann, T. 1995. Gibbon systematics and species identification. *International Zoo News, 42*, 467-501.

Geissmann, T. 2002a. Taxonomy and evolution of gibbons. In C. Soligo, G. Anzenberger, & R. D. Martin (Eds.), *Anthropology and primatology into the third millennium: The centenary congress of the Zurich anthropological institute* (pp. 28-31). Wiley-Liss: New York.

Geissmann, T. 2002b. Duet-splitting and the evolution of gibbon songs. *Biological Reviews, 77*, 57-76.

Geissmann, T. 2007. Status reassessment of the gibbons: Results of the Asian primate red list workshop 2006. *Gibbon Journal, 3*, 5-15.

Geissmann, T. 2009. Door slamming: Tool-use by a captive white-handed gibbon (*Hylobates lar*). *Gibbon Journal, 5*, 53-60.

Gibbon Conservation Center. 2018. Retrieved from https://www.gibboncenter.org/

Griffin, A. S. & Guez, D. 2014. Innovation and problem solving: A review of common mechanisms. *Behavioural Processes, 109*, 121-134.

Groves, C. P. 2001. *Primate Taxonomy*. Smithsonian Institution Press: Washington.

Harding, L. E. 2012. *Nomascus leucogenys* (Primates: Hylobatidae). *Mammalian Species, 44*, 1-15.

Hare, B. & Tomasello, M. 2004. Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. *Animal Behaviour, 68*, 571-581.

Harlow, H. F. 1932. Comparative behavior of primates. III. Complicated delayed reaction tests on primates. *Journal of Comparative Psychology, 14*, 241-252.

Harlow, H. F. 1949. The formation of learning sets. *Psychological Review, 56*, 51-65.

Harlow, H. F., Uehling, H. & Maslow, A. H. 1932. Comparative behavior of primates. I. Delayed reaction tests on primates from the lemur to the orang-outan. *Journal of Comparative Psychology, 13*, 313-343.

Harrison, T. 2016. The fossil record and evolutionary history of hylobatids. In U. H. Reichard, H. Hirai & C. Barelli (Eds.), *Evolution of gibbons and siamang* (pp. 91-110). Springer: New York.

Hauser, M. D. 1997. Artifactual kind and functional design features: What a primate understands without language. *Cognition, 64*, 285-308.

Hauser, M. D., Pearson, H. & Seelig, D. 2002. Ontogeny of tool use in cottontop tamarins, *Saguinus oedipus*: Innate recognition of functionally relevant features. *Animal Behaviour, 64*, 299-311.

Herrmann, E., Misch, A., Hernandez-Lloreda, V. & Tomasello, M. 2015. Uniquely human self-control begins at school age. *Developmental Science, 18*, 979-993.

Inoue, Y. & Inoue, E. 2002. The trap-table problem with a young white-handed gibbon (*Hylobates lar*). Retrieved from http://www.pri.kyotou.ac.jp/meetings/2002/saga5/abstract/poster1/yoichi.html

Irish, M. & Piguet, O. 2013. The pivotal role of semantic memory in remembering the past and imagining the future. *Frontiers in Behavioral Neuroscience, 7*, 27.

IUCN Red List. 2018. Retrieved from http://www.iucnredlist.org

Janmaat, K. R. L., Polansky, L., Ban, S. D. & Boesch, C. 2014. Wild chimpanzees plan their breakfast time, type, and location. *Proceedings of the National Academy of Sciences, 111*, 16343-16348.

Jarrard, L. E. & Moise, S. L. 1971. Short-term memory in the monkey. In L. E. Jarrard (Ed.), *Cognitive Processes of Nonhuman Primates* (pp. 1-24). Academic Press: New York.

Jerison, H. J. 1973. *Evolution of the Brain and Intelligence*. Academic Press: New York.

Kaufman, A. B., Reynolds, M. R. & Kaufman, A. S. 2019. The structure of ape (*Hominoidea*) intelligence. *Journal of Comparative Psychology, 133*, 92-105.

Kendal, R. L., Coe, R. L. & Laland, K. N. 2005. Age differences in neophilia, exploration, and innovation in family groups of callitrichid monkeys. *Amerian Journal of Primatology, 66*, 167-188.

Kim, S., Lappan, S. & Choe, J. C. 2011. Diet and ranging behavior of the endangered Javan gibbon (*Hylobates moloch*) in a submontane tropical rainforest. *American Journal of Primatology, 73*, 270-280.

Koda, H. 2016. Gibbon songs: Understanding the evolution and development of this unique form of vocal communication. In U. H. Reichard, H. Hirai & C. Barelli (Eds.), *Evolution of gibbons and siamang* (pp. 349-359). Springer: New York.

Köhler, W. 1925. *The mentality of Apes*. Harcourt, Brace & Company Inc.: New York.

MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., … Zhao, Y. 2014. The evolution of self-control. *Proceedings of the National Academy of Sciences, 111*, 2140-2148.

MacLeod, C. E., Zilles, K., Schleicher, A., Rilling, J. K. & Gibson, K. R. 2003. Expansion of the neocerebellum in Hominoidea. *Journal of Human Evolution, 44*, 401-429.

Malone, N. & Fuentes, A. 2009. The ecology and evolution of hylobatid communities: Causal and contextual factors underlying inter- and intraspecific variation. In S. Lappan & D. J. Whittaker (Eds.), *The gibbons: New perspectives on small ape socioecology and population biology* (pp. 241-264). Springer: New York.

Martin-Ordas, G. 2016. With the future in mind: Toward a comprehensive understanding of the evolution of future-oriented cognition. In K. Michaelian, S. B. Klein, & K. K. Szpunar (Eds.), *Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel* (pp. 306-327). Oxford University Press: Oxford.

Maslow, A. H. 1933. Comparative behavior of primates. VI. Food preferences of primates. *Journal of Comparative Psychology, 16*, 187-197.

Meldrum, R. C., Petkovsek, M. A., Boutwell, B. B. & Young, J. T. N. 2017. Reassessing the relationship between general intelligence and self-control in childhood. *Intelligence, 60*, 1-9.

Michilsens, F., Vereecke, E. E., D'Aout, K. & Aerts, P. 2009. Functional anatomy of the gibbon forelimb: Adaptations to a brachiating lifestyle. *Journal of Anatomy, 215*, 335-354.

Miloyan, B. & Suddendorf, T. 2015. Feelings of the future. *Trends in Cognitive Sciences, 19*, 196-200.

Mischel, W. & Metzner, R. 1962. Preference for delayed reward as a function of age, intelligence, and length of delay interval. *Journal of Abnormal and Social Psychology, 64*, 425-431.

Montgomery, S. H. 2014. The relationship between play, brain growth and behavioural flexibility in primates. *Animal Behaviour, 90*, 281-286.

Mootnick, A. R. 2006. Gibbon (*Hylobatidae*) species identification recommended for rescue or breeding centers. *Primate Conservation, 21*, 103-138.

Mootnick, A. R. & Fan, PF. 2011. A comparative study of crested gibbons (*Nomascus*). *American Journal of Primatology, 73*, 135-154.

Mootnick, A. R. & Groves, C. P. 2005. A new generic name for the hoolock gibbon (*Hylobatidae*). *International Journal of Primatology, 26*, 971-976.

Morino, L. 2009. Monogmay in mammals: Expanding the perspective on hylobatid mating systems. In S. Lappan & D. J. Whittaker (Eds.), *The gibbons: New perspectives on small ape socioecology and population biology* (pp. 279-311). Springer: New York.

Morris, R. G. M. 2001. Episodic-like memory in animals: Psychological criteria, neural mechanisms and the value of episodic-like tasks to investigate animal models of neurodegenerative disease. *Philosophical Transactions of the Royal Society of London B, 356*, 1453–1465.

Morton, F. B., Lee, P. C. & Buchanan-Smith, H. M. 2013. Taking personality selection bias seriously in animal cognition research: A case study in capuchin monkeys (*Sapajus apella*). *Animal Cognition, 16*, 677-684.

Mulcahy, N. J. & Call, J. 2006. Apes save tools for future use. *Science, 312*, 1038-1040.

Osvath, M. 2010. Great ape foresight is looking great. *Animal Cognition, 13*, 777-781.

Osvath, M. & Martin-Ordas, G. 2014. The future of future-oriented cognition in non-humans: Theory and the empirical case of the great apes. *Philosophical Transactions of the Royal Society of London B, 369*, 20130486.

Osvath, M. & Osvath, H. 2008. Chimpanzee (*Pan troglodytes*) and orangutan (*Pongo abelii*) forethought: Self-control and pre-experience in the face of future tool use. *Animal Cognition, 11*, 661-674.

Osvath, M. & Persson, T. 2013. Great apes can defer exchange: A replication with different results suggesting future oriented behavior. *Frontiers in Comparative Psychology, 4*, 1-8.

Osvath, M., Raby, C. R. & Clayton, N. S. 2009. What should be compared in comparative mental time travel? *Trends in Cognitive Sciences, 14*, 51-52.

Palombit, R. A. 1994. Extra-pair copulations in a monogamous ape. *Animal Behaviour, 47*, 721–723.

Parker, C. E. 1974. Behavioral diversity in ten species of nonhuman primates. *Journal of Comparative and Physiological Psychology, 87*, 930-937.

Paxton, R. & Hampton, R. R. 2009. Tests of planning and the Bischof-Köhler hypothesis in rhesus monkeys (*Macaca mulatta*). *Behavioural Processes, 80*, 238-246.

Penn, D. C. & Povinelli, D. J. 2007. Causal cognition in human and nonhuman animals: A comparative, critical review. *Annual Review of Psychology, 58*, 97-118.

Péter, A. 2018. Solomon Coder. Retrieved from https://solomoncoder.com/

Phoonjampa, R., Koenig, A., Borries, C. & Gale, G. A. 2010. Selection of sleeping trees in pileated gibbons (*Hylobates pileatus*). *American Journal of Primatology, 72*, 617-625.

Povinelli, D. J. 2000. *Folk Physics for Apes*. Oxford University Press: Oxford.

Povinelli, D. J. & Reaux, J. E. 2000. The trap-table problem. In D. J. Povinelli, *Folk Physics for Apes: The chimpanzee's theory of how the world works* (pp. 132-148). Oxford University Press: Oxford.

Povinelli, D. J. & Reaux, J. E. 2000. The inverted- and broken- rake problems. In D. J. Povinelli, *Folk Physics for Apes: The chimpanzee's theory of how the world works* (pp. 149-162). Oxford University Press: Oxford.

Preuschoft, H., Schönwasser, K. H. & Witzel, U. 2016. Selective value of characteristic size parameters in Hylobatids. A biomechanical approach to small ape size and morphology. In U. H. Reichard, H. Hirai & C. Barelli (Eds.), *Evolution of gibbons and siamang* (pp. 229-265). Springer: New York.

Prime, J. M. & Ford, S. M. Hand manipulation skills in Hylobatids. In U. H. Reichard, H. Hirai & C. Barelli (Eds.), *Evolution of gibbons and siamang* (pp. 269-289). Springer: New York.

Raby, C. R., Alexis, D. M., Dickinson, A. & Clayton, N. S. 2007. Planning for the future by western scrub-jays. *Nature, 445*, 919-921.

Raby, C. R. & Clayton, N. S. 2009. Prospective cognition in animals. *Behavioural Processes, 80*, 314-324.

Reader, S. M. & Laland, K. N. 2001. Primate innovation: Sex, age and social rank differences. *International Journal of Primatology, 22*, 787-805.

Reichard, U. H. & Barelli, C. 2008. Life history and reproductive strategies of Khao Yai *Hylobates lar*: Implications for social evolution in apes. *International Journal of Primatology, 29*, 823-844.

Reichard, U. H., Barelli, C., Hirai, H. & Nowak, M. G. 2016. The evolution of gibbons and siamang. In U. H. Reichard, H. Hirai & C. Barelli (Eds.), *Evolution of gibbons and siamang* (pp. 3-41). Springer: New York.

Reichard, U. H., Ganpanakngan, M. & Barelli, C. 2012. White-handed gibbons of Khao Yai: Social flexibility, complex reproductive strategies, and a slow life history. In P. M. Kappeler & D. Watts (Eds.), *Long-term field studies of primates* (pp. 237-258). Springer: Berlin.

Reichard, U. & Sommer, V. 1997. Group encounters in wild gibbons (*Hylobates lar*): Agonism, affiliation, and the concept of infanticide. *Behaviour, 134*, 1135-1174.

Rilling, J. K. 2006. Human and nonhuman primate brains: Are they allometrically scaled versions of the same design? *Evolutionary Anthropology, 15*, 65-77.

Rilling, J. K. & Insel, T. R. 1999. The primate neocortex in comparative perspective using magnetic resonance imaging. *Journal of Human Evolution, 37*, 191-223.

Roberts, W. A. 2007. Mental time travel: Animals anticipate the future. *Current Biology, 17*, 418-420.

Roberts, W. A. & Feeney, M. C. 2009a. The comparative study of mental time travel. *Trends in Cognitive Sciences, 13*, 271–277.

Roberts, W. A. & Feeney, M. C. 2009b. Temporal sequencing is essential to future planning: Response to Osvath, Raby and Clayton. *Trends in Cognitive Sciences, 14*, 52–53.

Rodriguez, J. S. & Paule, M. G. 2009. Working memory delayed response tasks in monkeys. In J. J. Buccafusco (Ed.), *Methods of Behavior Analysis in Neuroscience* (2nd ed.). CRC Press: Florida.

Roos, C. 2016. Phylogeny and classification of gibbons (*Hylobatidae*). In U. H. Reichard, H. Hirai & C. Barelli (Eds.), *Evolution of gibbons and siamang* (pp. 151-165). Springer: New York.

Roos, C. & Geissmann, T. 2001. Molecular phylogeny of the major hylobatid divisions. *Molecular Phylogenetics and Evolution, 19*, 486-494.

Rumbaugh, D. M. & McCormack, C. 1967. The learning skills of primates: A comparative study of apes and monkeys. In D. Stark, R. Schneider & H. J. Ruhn (Eds.), *Progressive primatology* (pp. 289-306). Lawerence Erlbaum Associates: New Jersey.

Santos, L. R., Miller, C. T. & Hauser, M. D. 2003. Representing tools: How two non-human primate species distinguish between the functionally relevant and irrelevant features of a tool. *Animal Cognition, 6*, 269-281.

Santos, L. R., Pearson, H. M., Spaepen, G. M., Tsao, F. & Hauser, M. D. 2006. Probing the limits of tool competence: Experiments with two non-tool-using species (*Cercopithecus aethiops* and *Saguinus oedipus*). *Animal Cognition, 9*, 94-109.

Sarma, K. & Kumar, A. 2016. The day range and home range of the eastern hoolock gibbon *Hoolock leuconedys* (*Mammalia: Primates: Hylobatidae*) in lower Dibang valley district in Arunachal Pradesh, India. *Journal of Threatened Taxa, 8*, 8641-8651.

Schacter, D. L., Benoit, R. G. & Szpunar, K. K. 2017. Episodic future thinking: Mechanisms and functions. *Current Opinion in Behavioral Sciences, 17*, 41-50.

Scheumann, M. & Call, J. 2006. Sumatran orangutans and a yellow-cheeked crested gibbon know what is where. *International Journal of Primatology, 27*, 575-602.

Schuppli, C., Forss, S., Meulman, E., Atmoko, S. U., van Noordwijk, M. & van Schaik, C. 2017. The effects of sociability on exploratory tendency and innovation repertoires in wild Sumatran and Bornean orangutans. *Scientific Reports, 7*, 15464.

Shifferman, E., M. 2009. Its own reward: Lessons to be drawn from the reversed-reward contingency paradigm. *Animal Cognition, 12*, 547–558.

Shultz, S. & Dunbar, R. I. M. 2010. Species differences in executive function correlate with hippocampus volume and neocortex ratio across nonhuman primates. *Journal of Comparative Psychology, 124*, 252-260.

Shumaker, R. W., Walkup, K. R. & Beck, B. B. 2011. *Animal Tool Behavior: The use and manufacture of tools by animals*. The John Hopkins University Press: Maryland.

Smith, R. E. 2003. The cost of remembering to remember in event-based prospective memory: Investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29*, 347-361.

Spaulding, B. & Hauser, M. 2005. What experience is required for acquiring tool competence? Experiments with two callitrichids. *Animal Behaviour, 70*, 517-526.

St. Amant, R. & Horton, T. E. 2008. Revisiting the definition of animal tool use. *Animal Behaviour, 75*, 1199-1208.

Suddendorf, T. 2010. Episodic memory versus episodic foresight: Similarities and differences. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 99-107.

Suddendorf, T. 2017. The emergence of episodic foresight and its consequences. *Child Development Perspectives, 11*, 191-195.

Suddendorf, T., Addis, D. R. & Corballis, M. C. 2009. Mental time travel and the shaping of the human mind. *Philosophical Transactions of the Royal Society of London B, 364*, 1317-1324.

Suddendorf, T., Bulley, A. & Miloyan, B. 2018. Prospection and natural selection. *Behavioral Sciences, 24*, 26-31.

Suddendorf, T. & Busby, J. 2003a. Mental time travel in animals? *Trends in Cognitive Sciences, 7*, 391-396.

Suddendorf, T. & Busby, J. 2005. Making decisions with the future in mind: Developmental and comparative identification of mental time travel. *Learning and Motivation, 36*, 110-125.

Suddendorf, T. & Corballis, M. C. 1997. Mental time travel and the evolution of the human mind. *Genetic Social and General Psychology Monographs, 123*, 133-167.

Suddendorf, T. & Corballis, M. C. 2007. The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences, 30*, 299-351.

Suddendorf, T. & Corballis, M. C. 2008. New evidence for animal foresight? *Animal Behaviour, 75*, 1-3.

Suddendorf, T. & Corballis, M. C. 2010. Behavioural evidence for mental time travel in nonhuman animals. *Behavioural Brain Research, 215*, 292-298.

Suddendorf, T., Corballis, M. C. & Collier-Baker, E. 2009. How great is great ape foresight? *Animal Cognition, 12*, 751-754.

Suwanvecho, U. & Brockelman, W. Y. 2012. Interspecific territoriality in gibbons (*Hylobates lar* and *H. pileatus*) and its effects on the dynamic of interspecies contact zones. *Primates, 53*, 97-108.

Szpunar, K. K., Spreng, R. N. & Schacter, D. L. 2016. Toward a taxonomy of future thinking. In K. Michaelian, S. B. Klein, & K. K. Szpunar (Eds.), *Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel* (pp. 21-38). Oxford University Press: Oxford

Takacs, Z., Morales, J. C., Geissmann, T. & Melnick, D. J. 2005. A complete species-level phylogeny of the hylobatidae based on mitochondrial *ND3-ND4* gene sequences. *Molecular Phylogenetics and Evolution, 36*, 456-467.

Tarullo, A. R., Obradović, J. & Gunnar, M. R. 2009. Self-control and the developing brain. *Zero to Three, 29*, 31-37.

Thinh, V. N., Mootnick, A. R., Geissmann, T., Li, M., Ziegler, T., Agil, M., Moisson, P., Nadler, T., Walter, L. & Roos, C. 2010a. Mitochondrial evidence for multiple radiations in the evolutionary history of small apes. *BMC Evolutionary Biology, 10*, 74-86.

Thinh, V. N., Mootnick, A. R., Thanh, V. N., Nadler, T. & Roos, C. 2010b. A new species of crested gibbon, from the central Annamite mountain range. *Vietnamese Journal of Primatology, 4*, 1-12.

Thinh, V. N., Rawson, B., Hallam, C., Kenyon, M., Nadler, T., Walter, L. & Roos, C. 2010c. Phylogeny and distribution of crested gibbons (genus *Nomascus*) based on mitochondrial cytochrome b gene sequence data. *American Journal of Primatology, 72*, 1047-1054.

Thom, J. M. & Clayton, N. S. 2016. Evolutionary perspectives on prospective cognition. In K. Michaelian, S. B. Klein, & K. K. Szpunar (Eds.), *Seeing the Future: Theoretical Perspectives on Future-Oriented Mental Time Travel* (pp. 287-305). Oxford University Press: Oxford.

Tomasello, M. & Call, J. 1997. *Primate Cognition*. Oxford University Press: New York.

Tulving, E. 1985. Memory and consciousness. *Canadian Psychology, 26*, 1-12.

Tulving, E. 2005. Episodic memory and autonoesis: Uniquely human? In H. S. Terrace, & J. Metcalfe (Eds.), *The Missing Link in Cognition: Origins of Self-Reflective Consciousness* (pp. 3-56). Oxford University Press: New York.

Tuttle, R. H. 1969. Quantitative and functional studies on the hands of the Anthropoidea I. The Hominoidea. *Journal of Morphology, 128*, 309-363.

Tuttle, R. H. 2014. *Apes and Human Evolution*. Harvard University Press: Massachusetts.

van Schaik, C. P., Damerius, L. & Isler, K. 2013. Wild orangutan males plan and communicate their travel direction one day in advance. *PLoS ONE, 8*, e74896.

Vlamings, P. H. J. M., Hare, B. & Call, J. 2010. Reaching around barriers: The performance of the great apes and 3-5-year-old children. *Animal Cognition, 13*, 273-285.

Whiten, A. & van de Waal, E. 2018. The pervasive role of social learning in primate lifetime development. *Behavioral Ecology and Sociobiology, 72*, 1-16.

Whittaker, D. J. & Lappan, S. 2009. The diversity of small apes and the importance of population-level studies. In S. Lappan & D. J. Whittaker (Eds.), *The gibbons: New perspectives on small ape socioecology and population biology* (pp. 3-10). Springer: New York.

Willemet, R. 2013. Reconsidering the evolution of brain, cognition and behavior in birds and mammals. *Frontiers in Psychology, 4*, 1-26.

Wood, D., Bruner, J. S. & Ross, G. The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*, 89-100.

Yerkes, R. M. & Yerkes, A. W. 1929. *The Great Apes: A study of anthropoid life.* Yale University Press: Connecticut.

Zhang, D., Fei, HL., Yuan, SD., Sun, WM., Ni, QY., Cui, LW. & Fan, PF. 2014. Ranging behavior of eastern hoolock gibbon (*Hoolock leuconedys*) in a northern montane forest in Gaoligongshan, Yunnan, China. *Primates, 55*, 239-247.